2015

Essays on High-Frequency Financial Data Analysis

Yingjie DONG
Singapore Management University, yj.dong.2010@phdecons.smu.edu.sg

Essays on High-Frequency Financial Data Analysis

YINGJIE DONG

SINGAPORE MANAGEMENT UNIVERSITY

2015

Essays on High-Frequency Financial Data Analysis

by
Yingjie Dong

Submitted to School of Economics in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Economics

Dissertation Committee:

Yiu-Kuen Tse (Supervisor/Chair)
Professor of Economics
Singapore Management University

Jun Yu
Professor of Economics and Finance
Singapore Management University

Anthony S Tay
Associate Professor of Economics
Singapore Management University

Aurobindo Ghosh
Assistant Professor of Finance
Singapore Management University

Singapore Management University
2015

Abstract

Essays on High-Frequency Financial Data Analysis

Yingjie Dong

This dissertation consists of three essays on high-frequency financial data analysis. I consider intraday periodicity adjustment and its effect on intraday volatility estimation, the Business Time Sampling (BTS) scheme and the estimation of market microstructure noise using NYSE tick-by-tick transaction data.

Chapter 2 studies two methods of adjusting for intraday periodicity of high-frequency financial data: the well-known Duration Adjustment (DA) method and the recently proposed Time Transformation (TT) method (Wu (2012)). I examine the effects of these adjustments on the estimation of intraday volatility using the Autoregressive Conditional Duration-Integrated Conditional Variance (ACD-ICV) method of Tse and Yang (2012). I find that daily volatility estimates are not sensitive to intraday periodicity adjustment. However, intraday volatility is found to have a weaker U-shaped volatility smile and a biased trough if intraday periodicity adjustment is not applied. In addition, adjustment taking account of trades with zero duration (multiple trades at the same time stamp) results in deeper intraday volatility smile.

Chapter 3 proposes a new method to implement the Business Time Sampling (BTS) scheme for high-frequency financial data using a time-transformation function. The sampled BTS returns have approximately equal volatility given a target average sampling frequency. My Monte Carlo results show that the Tripower Realized Volatility (TRV) estimates of daily volatility using the BTS returns produce smaller root mean-squared error than estimates using returns based on the Calendar

Time Sampling (CTS) and Tick Time Sampling (TTS) schemes, with and without subsampling. Based on the BTS methodology I propose a modified ACD-ICV estimate of intraday volatility and find that this new method has superior performance over the Realized Kernel estimate and the ACD-ICV estimate based on sampling by price events.

Chapter 4 proposes new methods to estimate the noise variance of high-frequency stock returns using differences of subsampling realized variance estimates at two or multiple time scales. Noise-variance estimates are compared and the new proposed estimates perform the best in reporting lower mean error and root mean-squared error. This chapter shows significant estimation error of noise-variance estimates when transactions are selected at too high or too low frequencies. For a typical New York Stock Exchange stock, the noise-to-signal ratio is around 0.005% in the period from 2010 to 2013.

Contents

Acknowledgements

I wish to express my deepest gratitude and thanks to my supervisor, professor Yiu Kuen TSE, for his excellent guidance and patience during my PhD study. I sincerely appreciate all his consistent contributions of time and ideas to me ever since year 2010. Without him my research could not have researched the current stage.

My special thank to professor Jun YU, professor Anthony S TAY and professor Aurobindo GHOSH, the committee members of my dissertation. Thanks for all of their brilliant comments and suggestions. I would also like to thank all professors in Singapore Management University for creating the stimulating research environments. Thanks to Lilian SEAH Wei Kim, Qiu Ling THOR, Lilian LIN and Esther PHANG Xin Yi, among others, for all of their help to me in SMU. Thanks to all of my best friends and they make my time in SMU so much more memorable.

Thanks to my beloved parents and my brother, Yaping DONG. They have been there day after day to make sure my life turned out this way. Thanks to my girlfriend, Yuxuan HUANG, for her years of love and support to me.

Chapter 1 Introduction

High-frequency financial data analysis has experienced vast and fast development over the past several years due to the rapid development of the theoretical methods and easier access to data. Applications of high-frequency data analysis are vast and important. For example, Andersen et al. (2013) point out that high-frequency data analysis provides a more accurate risk assessment in the financial risk measurement and risk management, treating both portfolio-level and asset-level analysis.

In high-frequency financial data analysis, we can model all transactions available or model a thinned point process which is defined by a subset of the transaction arrival times with specific characteristics. For example, by modeling the *transaction durations*, Engle and Russell (1998) estimate the arrival rate of traders using the Autoregressive Conditional Duration (ACD) model. In contrast, by modeling the *price durations*, Engle and Russell (1998) estimate the conditional instantaneous volatility using the ACD estimate and Tse and Yang (2012) estimate the conditional integrated volatility using the Autoregressive Conditional Duration-Integrated Conditional Variance (ACD-ICV) estimate. The price durations are calculated based on the price events which are obtained when the absolute price/logarithmic price change is bigger than or equal to the predetermined threshold value δ.

A well known problem in analyzing high-frequency financial data is the stylized fact of intraday periodicity: trading activities are usually higher at the beginning and close of the trading day than around lunch time, which is mainly due to opening auctions and closing effect separately. This trading pattern induces the average transaction duration and the average intraday volatility to exhibit an inverted U-shape over the trading day. Moreover, in Chapter 3, we provide empirical evidence of the difference of the intraday periodicity in trading activity from that in volatility.

To adjust for the intraday periodicity, we suggest using the *Time Transformation* (TT) method which is investigated in Wu (2012) and Chapter 2 and Chapter 3 in this dissertation. The idea of the TT method is that, with the help of a time-transformation function, we can switch all transactions between the *calendar time* and the *diurnally transformed time* easily. If the time-transformation function is constructed properly, all sampled transactions will be evenly spread out on the diurnally transformed time (the meaning of "evenly" will be different based on different economic issues). For example, in Chapter 2, to adjust for intraday periodicity in trading activity, we construct the time-transformation function using the number of transactions. As a result, all transactions on the diurnally transformed time will be evenly spread out over the trading day. In Chapter 3, to adjust for intraday volatility periodicity, we construct the time-transformation function using the estimated intraday integrated volatility. Thus all sampled durations with the same length on the diurnally transformed time will have approximately equal integrated volatility. Using the inverse of the time transformation function, we can obtain transactions with equal volatility increments at a pre-specified sampling frequency, which corresponds to the Business Time Sampling (BTS) scheme. There are many advantages of the TT method, as we mentioned in Chapter 2 and Chapter 3.

Applications of the TT method are wide and interesting. Specifically, in Chapter 2, we investigate the effects of intraday trading activity periodicity adjustments on the intraday volatility estimation using the ACD-ICV method of Tse and Yang (2012). One of a superior feature of the ACD-ICV estimate for the intraday volatility estimation is that we can make use of data outside the period of interest, based on an autoregressive price duration assumption. We find that daily volatility estimates are not sensitive to intraday periodicity adjustment. However, our findings tend to support the intraday periodicity adjustment for the intraday volatility estimation (eg. 30-min). In Chapter 3, using the TT method, we implement the Business Time

Sampling (BTS) scheme at high sampling frequency (eg. 3-min or 5-min). We then investigate two applications of the BTS scheme: testing whether a logarithmic price process is a time-changed Brownian motion after correcting for the drift and price jumps (which is related to the no-arbitrage condition) and estimating the integrated volatility using the high-frequency *BTS returns* or *BTS durations*. We find that the log-price process of the investigated NYSE stocks can be considered as a time-changed Brownian motion with drift and price jumps and the Gaussian distribution assumption describes the BTS returns better than returns obtained from the Calendar Time Sampling (CTS) scheme and Tick Time Sampling (TTS) scheme. Due to better iid Gaussian property, BTS transactions performs better than CTS and TTS transactions in estimating intraday integrated volatility by reporting smaller root mean-squared error (RMSE). This observation is interesting since we lose around half of observations when implementing the BTS scheme with subsampling. We also propose a modified ACD-ICV estimate by modeling the BTS durations and show its superiority over the Realized Kernel estimate and the ACD-ICV estimate by modeling price durations.

In high-frequency data analysis, we also have the market microstructure noise problem due to the presence of the bid ask spread, price discreteness and asymmetric information of traders, among other reasons. Wide literature are developed to investigate the noise properties and the noise effect in High-frequency financial data analysis. In Chapter 4, we show the negative cross-correlation between the noise and latent returns. We propose new noise-variance estimates by using difference of subsampling realized variance estimates. We compare different noise-variance estimates in our Monte Carlo simulation study and find that our proposed noise-variance estimates performs the best by reporting the mean error (ME) and root mean-squared error (RMSE) most closely to zero. Empirically we find the noise-to-signal ratio (NSR) is around 0.005% for the NYSE stocks in recent years. This is largely smaller in compare with magnitude estimated in previous literature. For example, Hansen and Lunde (2006) report that the NSR is below 0.1% for DJIA stocks in year 2000. The magnitude of microstructure noise variance provide very helpful information to high-frequency traders and market regulators. For example, Hasbrouck (1993) points out that it is a good measure of market quality and can be used to evaluate brokers performance or to determine markets or regulatory structures under which transaction costs are minimized. The small magnitude of the NSR suggests that it is important to consider the cross-correlation between the noise and latent returns when researchers investigate the noise properties.

Chapter 2 Intraday Periodicity Adjustments of Transaction Duration and Their Effects on High-frequency Volatility Estimation

2.1 Introduction

A well known problem in analyzing high-frequency financial data is the stylized fact of intraday periodicity: trading activities are usually higher at the beginning and close of the trading day than around lunch time. This trading pattern induces the average transaction duration to exhibit an inverted U-shape over the trading day. Andersen and Bollerslev (1997) point out that volatility over different intervals of the same calendar-time length at different times of the daymay differ due to their differences in trading activities.

Recognizing the empirical fact of duration clustering, Engle and Russell (1998) introduce the Autoregressive Conditional Duration (ACD) model. They propose to correct for the intraday duration periodicity prior to fitting the ACD model to the data. Specifically, they apply the Duration Adjustment (DA) method to adjust for transaction duration as

$$\tilde{x}_{i+1} = \frac{x_{i+1}}{\phi(t_i)}, \tag{2.1.1}$$

where t_i is the calendar time of occurrence of the ith trade, $x_{i+1} = t_{i+1} - t_i$ is the duration of the $(i+1)$th trade in calendar time, $\tilde{x}_{i+1}$ is the diurnally adjusted duration of the $(i+1)$th trade and $\phi(\cdot)$ is the diurnal adjustment factor with its argument usually taken as the calendar time t_i.

Recently Wu (2012) proposes a new method to correct for intraday periodicity, which will be called the Time Transformation (TT) method in this paper. The theoretical underpinning of the TT method is that if there were no intraday differences in trading activities, we would expect the transactions to be evenly spread out throughout the trading day. Under the TT method a time-transformation function is determined using empirical data so that the transactions are evenly observed throughout the day under the transformed time.

To assess the empirical implications of these adjustment methods, we study their effects on daily and intraday volatility estimations. Following the method proposed by Tse and Yang (2012) for the estimation of high-frequency volatility, called the Autoregressive Conditional Duration-Integrated Conditional Variance (ACD-ICV) method, we examine the effects of the use of the DA and TT methods to adjust for intraday transaction durations on the estimation of intraday volatility. Our findings are as follows. First, for the estimation of daily volatility, whether or not the durations are diurnally adjusted makes little difference. Second, to estimate intraday volatility, correcting for intraday periodicity using either the DA or TT method produces more prominent U-shaped volatility smiles than not adjusting for intraday periodicity. Third, deeper intraday volatility smile is observed if the duration adjustment treats multiple trades with the same time stamp as separate trades (i.e., data with zero duration are retained). Fourth, intraday volatility smile has a biased trough if intraday periodicity adjustment is not applied. The balance of this paper is as follows. Section 2 outlines the DA and TT methods. Section 3 summarizes the ACD-ICV method for estimating intra-day volatility. Section 4 reports the empirical results. Section 5 concludes.

2.2 Diurnal adjustment for intraday periodicity

2.2.1 The Duration Adjustment (DA) Method

The DA method involves adjusting the raw duration using a diurnal factor to obtain the diurnally adjusted duration, as given in Eq. (1). To specify the diurnal factor $\phi(\cdot)$ many studies use the regression method with linear spline or cubic spline method proposed by Engle and Russell (1998).While there are quite a few variations in the literature in the estimation of $\phi(\cdot)$, we follow the method proposed by Bauwens and Giot (2000), for which the focus is on estimating the expected duration conditional on the time of the day. The details of the procedure are summarized in Appendix A.1.

The DA method has some important drawbacks (see also Wu (2012)). First, the average duration in each interval is a local measure. There is a choice between taking longer intervals (so that there are more observations in each interval) versus shorter intervals (so that the knots are more precisely located). Second, given a raw duration, it is not clear which time point within the interval should be taken to evaluate the diurnal factor. As the diurnal-factor function takes an inverted U-shape, if the start time of the interval is taken for adjustment, which is the usual practice, the adjustment may be understated for trades before lunch time and overstated for trades after lunch time.[1] Hence, a systematic bias may be introduced in the diurnal adjustment.

2.2.2 The Time Transformation (TT) Method

We now briefly outline the TT method. Let t denote the calendar time and $\tilde{t}$ denote the corresponding diurnally transformed time (in seconds from the begin-

[1]The phenomenon of an inverted U-shape diurnal-factor function may be due to market participants' different trading purposes at different periods. Market participants tend to trade heavily when the market is open due to the opening auctions. They trade more heavily near the end of the day due to the closing effect, closing their positions before the market closes.

ning of trade). In the context of the New York Stock Exchange (NYSE), as one trading day has 6.5 h (23,400 s), $0 \leq t \leq 23,400$. We denote $Q(t)$ as the empirical average proportion of trades in a day up to time t and define the diurnally transformed time $\tilde{(}t) = 23,400Q(t)$. Conversely, given a diurnally transformed time $\tilde{(}t)$, the corresponding calendar time is $t = Q^{-1}(\tilde{(}t)/23,400)$, where $Q^{-1}(\cdot)$ is the inverse function of $Q(\cdot)$. Appendix A.2 provides more details for the computation of $Q(\cdot)$.

Given any two calendar-time points $t_i < t_j$, the diurnally adjusted duration between these two time points $\tilde{t}_j - \tilde{t}_i = 23,400[Q(t_j) - Q(t_i)]$. Likewise, given any two diurnally transformed time points $\tilde{t}_i < \tilde{t}_j$, the corresponding duration in calendar time is

$$Q^{-1}\left(\frac{\tilde{t}_j}{23400}\right) - Q^{-1}\left(\frac{\tilde{t}_i}{23400}\right). \tag{2.2.1}$$

There are some advantages of the TT method over the DA method. First, the $Q(\cdot)$ function is easy to compute and it depends on all data in the sample. Second, the definition of diurnally adjusted duration is natural. This removes the ambiguity in the choice of the time point within the transaction interval for the application of the diurnal factor. Third, the switch between calendar time and diurnally adjusted time can be performed easily. This facilitates simulation using models in one measure of duration (e.g., ACD model for diurnally adjusted duration) to draw implications for the market in another measure of time (e.g., implications for the market in calendar time). Recently, Dionne et al. (2009) suggest using the simulation method to estimate the intraday value at risk (IVaR).While the time interval specified may be in calendar time, the duration model estimated may be for diurnally adjusted data. The TT method will be convenient to use as the calendar time and diurnally adjusted time can be easily converted from one another.

2.3 Diurnal adjustment and intraday volatility estimation

To examine the empirical implications of the methods of adjusting for intraday periodicity we consider the effects of these adjustments on the estimation of intraday volatility.[2] We use the ACD-ICV method for estimating high-frequency volatility proposed by Tse and Yang (2012), and compare the volatility estimates when the data are (a) not adjusted for intraday periodicity, (b) adjusted for intraday periodicity using the TT method, and (c) adjusted for intraday periodicity using the DA method. In what follows we first briefly summarize the ACD-ICV method, after which we report the empirical results on the NYSE data.

2.3.1 The ACD-ICV Method

The ACD-ICV method samples observations from transaction data based on a threshold price range δ. A price event is said to occur if the logarithmic stock price first moves by an amount δ or more, whether upwards or downwards. The waiting time for the price event to occur is called the price duration. Let t_0 be the beginning time of a period, and $t_1, t_2, \cdots, t_N$ be subsequent times of occurrence of price events, so that the price duration of the ith trade is $x_i = t_i - t_{i-1}$. The integrated conditional variance (ICV) over the interval (t_0, t_N), which may be one trading day or a subinterval of a trading day, is given by

$$\mathrm{ICV} = \delta^2 \sum_{i=0}^{N-1} \frac{t_{i+1} - t_i}{\psi_{i+1}}, \qquad (2.3.1)$$

where $\varepsilon_i = x_i/\psi_i$ are assumed to be i.i.d. standard exponential. Estimates of ψ_i are then substituted into Eq. (3), resulting in the ACD-ICV estimate of the vari-

[2]Research on intraday movements of equity prices has recently attracted much interest due to its implications for market participants such as day traders and market makers, as illustrated by the works of Giot (2005), Dionne et al. (2009), and Liu and Tse (2013).

ance in the interval (t_0, t_N). In this paper we use the power ACD (PACD) model of Fernandes and Grammig (2006) to estimate the conditional duration.

2.3.2 Three Intraday Volatility Estimates

We consider three alternative ways of estimating daily or intraday volatility using the ACD-ICV method. First, we compute the price duration using unadjusted calendar time, and call this the raw duration. We estimate the PACD model using the raw duration, with resulting estimates of ψ_i denoted by $\hat{\psi}_i$ and the ICV computed as

$$\text{ICV} = \delta^2 \sum_{i=0}^{N-1} \frac{x_{i+1}}{\hat{\psi}_{i+1}} = \delta^2 \sum_{i=0}^{N-1} \frac{t_{i+1} - t_i}{\hat{\psi}_{i+1}}. \tag{2.3.2}$$

We call this method M1, which uses raw duration without diurnal adjustment. Next, we take account of intraday periodicity using the TT method and compute the diurnalized duration as

$$\tilde{x}_{i+1} = \tilde{t}_{i+1} - \tilde{t}_i. \tag{2.3.3}$$

We estimate the PACD model using the TT diurnalized duration, with resulting estimates of ψ_i denoted by $\hat{\psi}_i$ and the ICV computed as

$$\text{ICV} = \delta^2 \sum_{i=0}^{N-1} \frac{\tilde{x}_{i+1}}{\tilde{\psi}_{i+1}} = \delta^2 \sum_{i=0}^{N-1} \frac{\tilde{t}_{i+1} - \tilde{t}_i}{\tilde{\psi}_{i+1}}. \tag{2.3.4}$$

We call this method M2, which uses diurnalized duration based on the TT method. Finally, we denote the diurnally adjusted duration using the DA method by $\breve{x}_i$, which is given by

$$\breve{x}_{i+1} = \frac{t_{i+1} - t_i}{\phi(t_i^M)} = \frac{x_{i+1}}{\phi(t_i^M)} \tag{2.3.5}$$

where $t_i^M = (t_i + t_{i+1})/2$ is the mid-point of the interval (t_i, t_{i+1}).[3] We estimate the PACD model using the diurnalized duration computed from the DA method, with

[3] Most studies in the literature evaluate the diurnal factor $\phi(\cdot)$ at the starting point ti of the interval. As our average price duration is around 5 min, it may be more appropriate to take the mid-point.We will report the robustness of our empirical results when the starting value of the interval is used to evaluate the diurnal factor instead.

resulting estimates of ψ_i denoted by $\breve{\psi}_i$ and the ICV computed as

$$\mathrm{ICV} = \delta^2 \sum_{i=0}^{N-1} \frac{\breve{x}_{i+1}}{\breve{\psi}_{i+1}} = \delta^2 \sum_{i=0}^{N-1} \frac{t_{i+1} - t_i}{\breve{\psi}_{i+1}\phi(t_i^M)}. \tag{2.3.6}$$

We call this method M3, which uses diurnalized duration based on the DA method.

2.4 Empirical results

The transaction data used in this paper are extracted and compiled from the NYSE Trade and Quote (TAQ) Database provided through the Wharton Research Data Services. We select ten stocks from the component stocks of the S&P500 index. Data from the Consolidated Trade (CT) file, including the date, trading time, price and number of shares traded are extracted for each stock, over the period January 3, 2005 through December 31, 2007. We drop some abnormal trading days that may contaminate the results. These include trading days with first trade after 11:00 or last trade before 13:30. The data are then filtered before the empirical analysis.We delete entries with corrected trades, i.e., trades with a correction indicator. Only data with a regular sale condition and code E and F are selected.[4]

2.4.1 Description of the NYSE Transaction Data

Table 2.1 presents some summary statistics of the selected stocks. Trades with the same time stamp can be treated as separate trades or one trade, and we report summary statistics for both assumptions.[5] It can be seen that all stocks in our sample are very actively traded. The number of transactions per day ranges from 4746 to 8264 and the average duration per trade ranges from 2.83 s to 4.93 s, when multiple trades are treated as one trade. On the other hand, when multiple trades are treated

[4]These screening procedures were adopted from Barndorff-Nielsen et al. (2009).

[5]When there are multiple trades at the same stamped time, we calculate the price at the stamped time as the average of the prices weighted by trade volume.

as separate trades, the average number of transactions per day increases substantially and ranges from 7360 to 14,591, while the average duration per trade ranges from 1.6 s to 3.18 s. The LjungCBox statistics strongly support serial correlation in durations, whether or not they are adjusted for intraday periodicity (using the TT method).[6] Nonetheless, the LjungCBox statistics are smaller for TT adjusted durations versus raw durations, and for data without zero durations versus with zero durations.

Figure 2.1 plots the total number of trades of IBMat each second from 9:30 to 16:00 over all trading days in the sample, withmultiple trades treated as separate trades. It can be seen that there is significant intraday periodicity: the number of trades at each second has a clear U-shape. Figure 2.2 plots the average trade durations from9:30 to 16:00,with smoothing performed using cubic spline. It can be seen that the average trade duration as a function of calendar time exhibits an inverted U-shape.[7] Figure 2.3 presents the plots of the diurnally transformed time $\tilde{t}$ against the calendar time t for the IBMstock. The solid line is the plot of 23,400$Q(t)$, computed by treatingmultiple trades as separate trades. Thus, if there was no intraday periodicity, 23,400$Q(t)$ should follow the 45 degree line.[8] Note that the $Q(t)$ function constructed is strictly monotonically increasing, so that t and $\tilde{t}$ form a one-to-one correspondence.

Figure 2.4 plots the smoothed average duration by time of the day using the TT adjusted data. There is little intraday variation in the average duration (note the scales on the vertical axis), which shows that intraday periodicity has been largely removed. We plot the TT adjusted durations over a selection of three days from the sample period in Figure 2.5. It can be seen that there is still duration clustering

[6]This is in contrast to Wu (2012) who cleaned the data to get rid of zero durations. Also, Wu (2012) used a shorter sample for their empirical study.

[7]To save space, only the graphs for IBM are presented. Results for other stocks are very similar, and they are not presented here.

[8]The vertical axis in Figure 2.3 is converted from second to transformed time from 9:30 to 16:00 for ease of reference.

for the data, suggesting that intraday periodicity and duration clustering are two separate phenomena. While TT adjustment can get rid of intraday periodicity, the adjusted data are still subject to duration clustering, possibly due to the lag effects in absorbing market news. This supports the modeling of the (adjusted and unadjusted) durations using the ACD models.

2.4.2 Empirical Estimates of Daily and Intraday Volatility

We compute the daily volatility estimates of the 10 selected stocks in our sample using the three methods: M1, M2 and M3.We report in detail the case when M2 is based on the convention that multiple trades are treated as separate trades (i.e., with zero durations) and M3 is based on the convention that multiple trades are treated as one trade (i.e., without zero durations). Table 2.2 reports the results for the daily volatility estimates, giving themean of each estimate, and themean absolute deviation (MAD) and the root mean-squared deviation (RMSD) between pairs of daily estimates. The daily volatility estimates are expressed in annualized return standard deviation in percentage. The results show that the daily volatility estimates of the three methods are very similar. In terms of the MAD, the differences between M2 and M3 (i.e., two methods with adjustment for intraday periodicity) are very small (less than 0.1% for all stocks except CVX). The differences betweenM1 versusM2 andM1 versusM3are, however, slightly higher, but still less than 0.22% for all stocks. The results based on RMSD give similar conclusion.

To investigate the effects of diurnal adjustment on intraday volatility we consider half-hour intervals from 9:30 through 16:00. Figure 2.6 plots the average intraday volatility estimates over all trading days for all 10 stocks, with the mid-point of each half-hour interval identified in the x-axis.[9] It can be seen that M2 and M3, which correct for intraday periodicity, are quite close to each other. However, the volatility estimated for the last half-hour is higher for the TT method (M2) than the

[9]To the best of our knowledge, this is the first documentation of intraday volatility estimates over half-hour intervals.

DA method (M3) for all stocks. Both M2 and M3 have more prominent intraday volatility smile (a deeper U-shape) than the volatility estimate with no periodicity correction (M1). Furthermore, while the lowest intraday volatility for M2 and M3 occurs between 12:45 and 13:15, the lowest intraday volatility for M1 occurs at 13:45 for almost all stocks. As the average duration curve in Figure 2.2 peaks around 13:00, the intraday volatility curve using M1 appears to be biased.[10] These findings have important implications for research on IVaR.

To examine the robustness of the results, we vary the treatment of multiple trades for M2 and M3, as well as the evaluation of the diurnal factor $\phi(\cdot)$. First, when the same treatment of multiple trades is applied to the two periodicity adjustment methods, their results are very similar and almost undistinguishable. Thus, the difference in the average volatility at the close of the market is due to the treatment of multiple trades rather than the choice of the method of adjustment. Second, if we evaluate $\phi(\cdot)$ in M3 at the starting point of the interval rather than the mid-point, M2 is found to have a slightly deeper volatility smile than M3, even when the same treatment of multiple trades is applied. Thus, M3 is not insensitive to the choice of the time point of correction. This is an important drawback of the M3 method and is generally overlooked in the literature.

2.5 Conclusion

In this paper we investigate the use of two methods in dealing with intraday periodicity of high-frequency stock price data, the Duration Adjustment (DA) method and the Time Transformation (TT) method, for the estimation of intraday volatility. The DA method diurnally adjusts the raw durations by estimating a diurnal factor $\phi(\cdot)$, while the TT method transforms the calendar time to the diurnally adjusted time from which adjusted duration is calculated.

Our empirical results show that whether or not the duration data are adjusted

[10]This bias is observed in all 10 stocks in our study.

for intraday periodicity has little impact on the daily volatility estimates using the ACD-ICV method. However, we document a clear regular intraday volatility smile, which shows a more prominent U-shape when the duration data are corrected for intraday periodicity than when it is not. An important issue in the implementation of the adjustment is how multiple trades at the same time stamp are treated. Intraday volatility smiles are more prominent if multiple trades are treated as separate trades rather than one trade, regardless of the method of adjustment. Thus, deleting observations with zero duration dampens the effects of intraday periodicity and may lead to under-estimated intraday volatility in periods of high trading activities.

While the DA method is widely used in the literature, the TT method has several clear advantages. The theoretical motivation of the TT method is quite intuitive and its empirical implementation is straightforward. Also, the results of the DA method are not robust to the point of evaluation of the diurnal factor. For studies that require simulating duration data, such as for the estimation of intraday value at risk, the TT method is simple to use.

Table 2.1: Summary statistics of stocks

	CVX	GE	IBM	JNJ	JPM	PFE	PG	T	WMT	XOM
Number of days	748	748	748	747	747	747	748	740	748	748
Multiple trades at the same time treated as separate trades										
Number of transactions (million)	7.709	7.107	5.919	5.498	6.544	6.305	5.598	5.647	6.280	10.914
Avg transactions per day	10306	9501	7913	7360	8760	8440	7484	7631	8396	14591
Avg duration (second)	2.27	2.46	2.96	3.18	2.67	2.77	3.13	3.07	2.79	1.60
Multiple trades at the same time treated as one trade										
Number of transactions (million)	4.898	4.492	4.006	3.807	4.109	4.352	3.721	3.512	4.289	6.181
Avg transactions per day	6548	6005	5356	5096	5501	5826	4975	4746	5735	8264
Avg duration (second)	3.57	3.90	4.37	4.59	4.25	4.02	4.70	4.93	4.08	2.83

Table 2.2: Results of daily volatility estimation

	Mean Daily ICV			MAD			RMSD		
Stock	M1	M2	M3	M1-M2	M1-M3	M2-M3	M1-M2	M1-M3	M2-M3
CVX	19.1913	19.2013	19.2085	0.1845	0.0969	0.1013	0.2513	0.1296	0.1389
GE	12.3780	12.3861	12.3822	0.1159	0.0720	0.0483	0.1722	0.1126	0.0691
IBM	14.7162	14.7506	14.7408	0.1058	0.0787	0.0376	0.1382	0.1069	0.0497
JNJ	11.1155	11.1476	11.1371	0.0839	0.0618	0.0286	0.1080	0.0800	0.0377
JPM	16.3696	16.3481	16.3473	0.2148	0.1562	0.0635	0.2988	0.2224	0.0863
PFE	16.1752	16.2178	16.2047	0.0962	0.0778	0.0295	0.1181	0.0966	0.0373
PG	12.8189	12.8722	12.8547	0.1021	0.0799	0.0313	0.1275	0.1014	0.0405
T	16.0181	16.0427	16.0366	0.1518	0.0963	0.0712	0.2302	0.1522	0.1025
WMT	14.9632	15.0001	14.9904	0.0894	0.0629	0.0389	0.1146	0.0813	0.0498
XOM	17.9755	17.9809	17.9752	0.1744	0.0881	0.0923	0.2236	0.1158	0.1170

Notes: Volatility is annualized standard deviation in percentage. MAD is mean absolute deviation. RMSD is Root mean-squared deviation. M1 uses raw duration without diurnal adjustment. M2 uses diurnalized duration based on the TT method. M3 uses diurnalized duration based on the DA method.

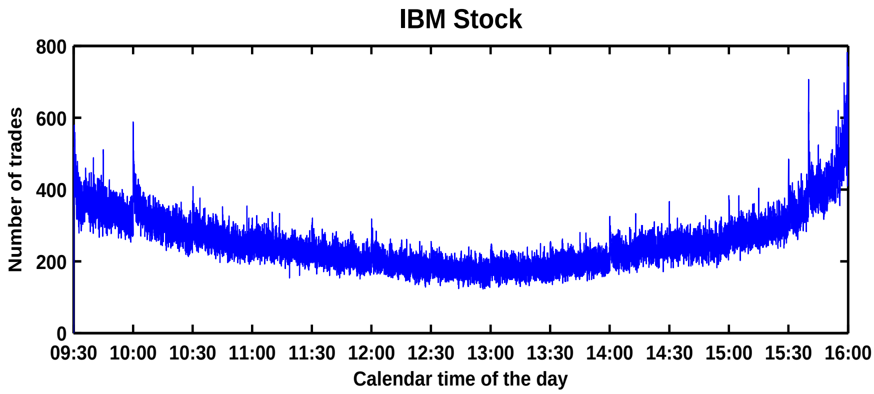

Figure 2.1: Number of trades at each second over all trading days.

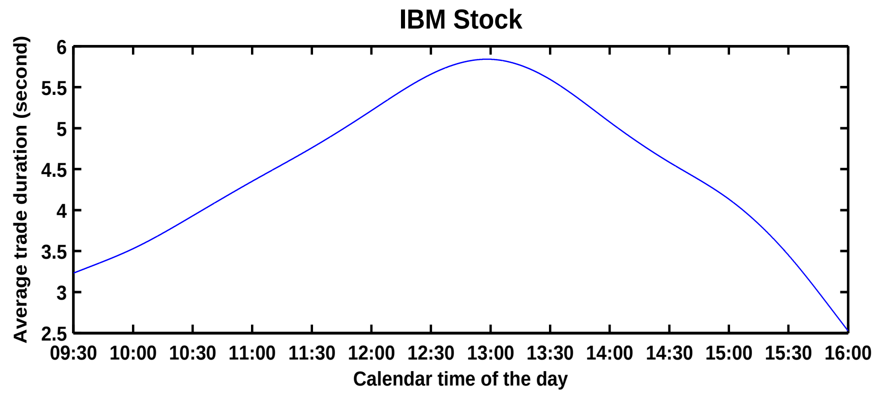

Figure 2.2: Smoothed average duration by time of the day over all trading days.

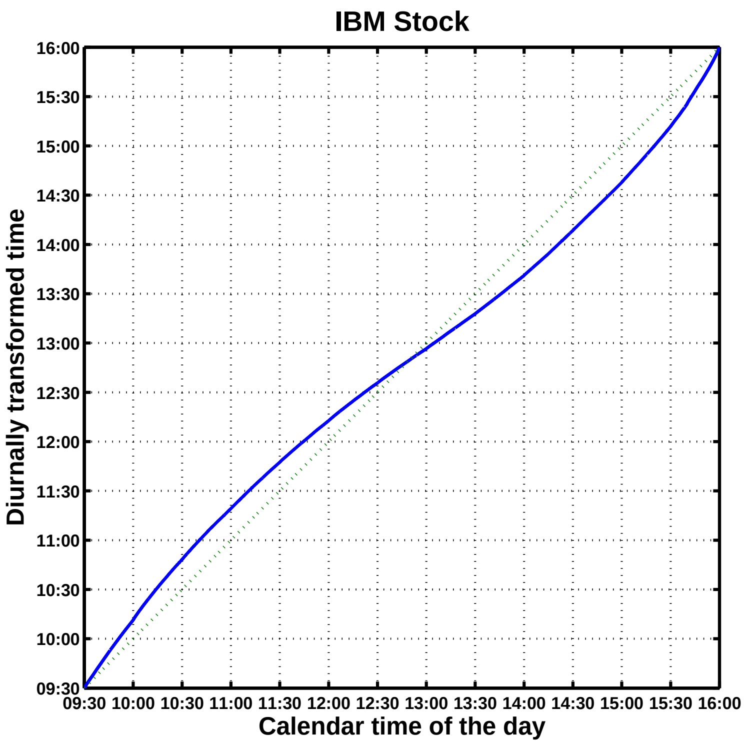

Figure 2.3: Diurnally transformed time versus calendar time.

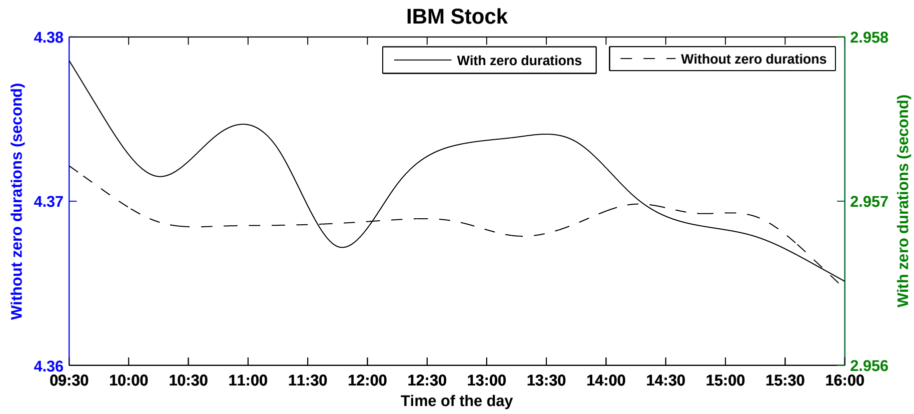

Figure 2.4: Smoothed average TT adjusted duration by time of the day over all trading days.

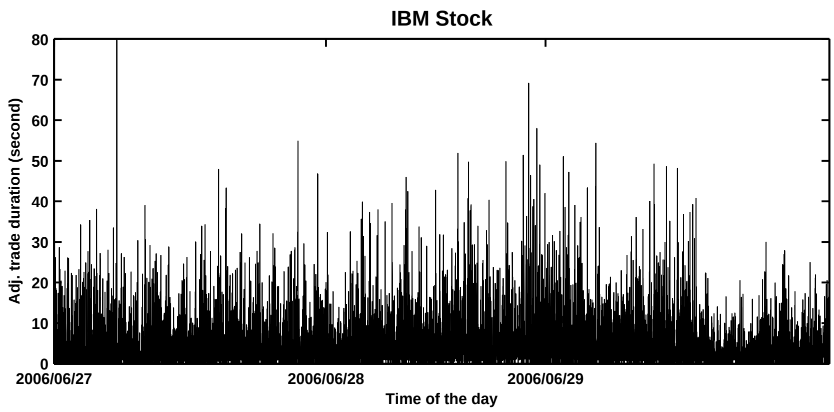

Figure 2.5: TT adjusted durations over 3 days in sample period.

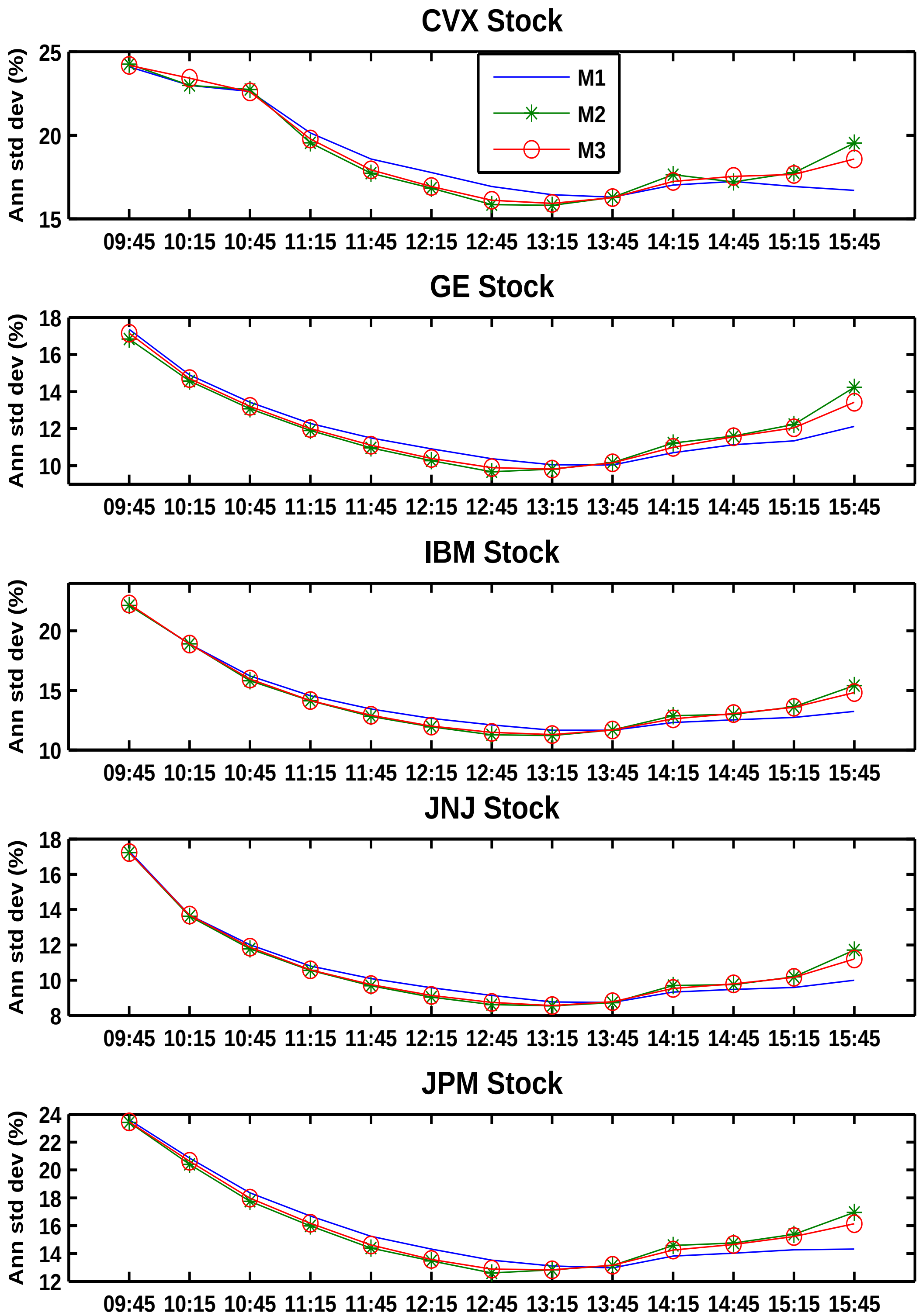

Figure 2.6: Intraday volatility calculated using M1, M2 and M3.

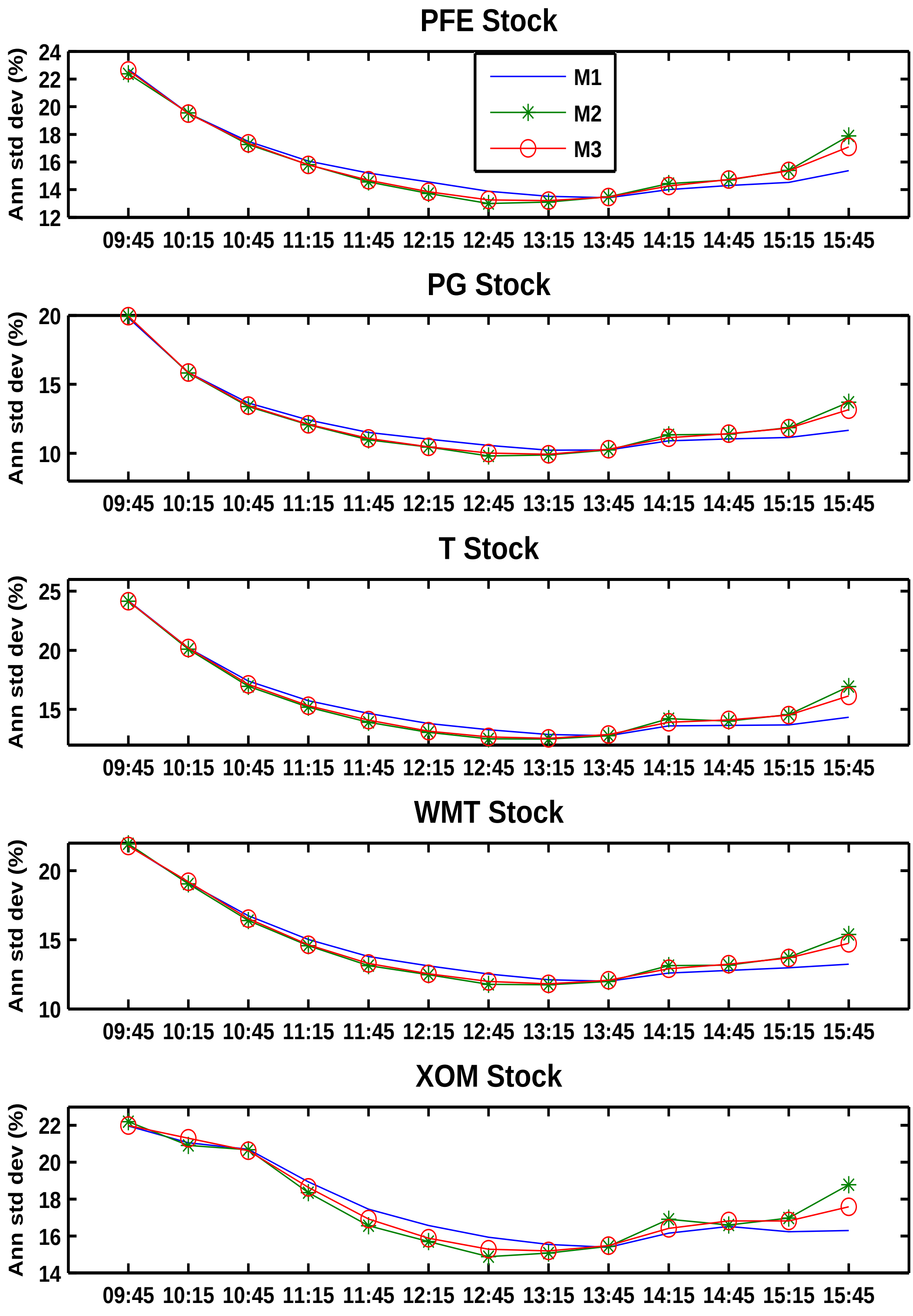

Figure 2.6 (continue): Intraday volatility calculated using M1, M2 and M3.

Chapter 3 Business Time Sampling Scheme and Its Application

3.1 Introduction

In high-frequency financial data analysis, researchers usually do not use all available data but would select a subgrid of transactions. To choose the subgrid, two issues have to be considered: selecting the sampling scheme and choosing the target average sampling frequency. Three sampling schemes are commonly used in the literature: Calendar Time Sampling (CTS), Tick Time Sampling (TTS) and Business Time Sampling (BTS). Under the CTS scheme, transactions are selected by regularly spaced calendar time, such as every 5 sec/min. The TTS scheme selects transactions with regularly spaced number of ticks, e.g., every 5 or 10 ticks. The BTS transactions are often selected to ensure approximately equal volatility for the returns over each interval. Thus, the CTS and TTS schemes are implemented based on explicit criteria (i.e., regular calendar-time length or number of ticks, respectively). In contrast, the BTS scheme depends on the unobserved volatility. As a result, the CTS and TTS schemes have been used widely in the literature, while the BTS scheme is used less frequently.

The BTS scheme possesses some desirable properties for high-frequency financial data analysis. In particular, it yields independently and identically distributed (iid) normal returns for a semimartingale price process even when there is leverage ffect.[1] In contrast, the calendar-time returns may not be iid normal even if the

[1] Dambis (1965) and Dubins and Schwartz (1965) show that a process compiled from a continu-

price process is a continuous local martingale.[2] The assumption of iid Gaussian returns is required for the Gaussian-likelihood approach (Nowman (1997)) and several widely used integrated volatility estimates, including the multipower variation (MPV) estimate of Barndorff-Nielsen *et al.* (2006), the quantile realized volatility (QRV) method of Christensen *et al.* (2010) and the nearest neighbor truncation method of Andersen *et al.* (2011) and Andersen *et al.* (2012). Thus, the importance of sampling returns that are iid Gaussian cannot be over-emphasized. Furthermore, the BTS scheme is useful for thinning the empirical price process, mitigating the effects of market microstructure noise.

The BTS scheme dates back to Dacorogna *et al.* (1993), Zhou (1998), and Peters and de Vilder (2006). Andersen *et al.* (2007) and Andersen *et al.* (2010) find that the normalized daily and weekly returns sampled in financial time, defined by equidistant increments of the realized volatility, accord well with the standard normal distribution. To sample the time points, they use a sequential method to include intraday returns until the cumulative squared returns exceeds the average daily or weekly realized volatility. In this paper we propose a new method to implement the BTS scheme. We use a time-transformation function to obtain the BTS time points, where the time-transformation function is computed based on intraday integrated volatility estimates.

We test the semi-martingale hypothesis on the BTS returns of 40 stocks selected from the New York Stock Exchange (NYSE). We further explore the use of the BTS scheme in estimating intraday volatility. First, we consider the Realized Volatility (RV) estimates for daily integrated volatility when the returns are sampled using the BTS, CTS and TTS schemes. Using the Tripower Realized Volatility (TRV) estimates of Barndorff-Nielsen *et al.* (2006), our Monte Carlo simulation shows that the TRV estimates (with and without subsampling) using the BTS returns provide the smallest root mean-squared error (RMSE). Second, we modify the Autoregressive

---

ous local martingale with equal quadratic-variation increments is a Brownian motion.

[2]This may be due to the leverage effect or varying volatility.

Conditional Duration-Integrated Conditional Volatility (ACD-ICV) method of Tse and Yang (2012), making use of the BTS scheme. Our modified ACD-ICV estimator performs better than the Realized Kernel (RK) estimates (Barndorff-Nielsen *et al.* (2008)) and the method of Tse and Yang (2012) based on price-event sampling.

The rest of this paper is as follows. Section 2 outlines our proposed implementation of the BTS scheme. Section 3 reports some empirical results on testing the semi-martingale hypothesis. In Section 4 we outline the estimation methods of daily volatility examined in this paper. We report the results of our Monte Carlo study in Section 5 and draw conclusions in Section 6. The Appendix provides further details of the jump detection procedure and computation of the BT time transformation function. Some additional results can be found in the accompanying supplementary material.

3.2 Intraday Periodicity and the BTS Scheme

3.2.1 Intraday Periodicity Patterns in Volatility and Trading Activity

To adjust for the intraday periodicity of transaction activity, Tse and Dong (2014) use a time transformation function computed by pooling all transactions over all trading days in the sample. The time-transformation function $Q(t)$ at calendar time t (in sec) of the trading day is computed as the empirical proportion of the number of trades up to time t over all trading days.[3] The diurnally transformed time $\tilde{t}$ for the calendar-time point t is defined by $\tilde{t} = 23400Q(t)$ and the diurnally transformed duration for the period between calendar time points t_i and $t_j(t_j > t_i)$ is $t_j > t_i$. In this paper we adopt this approach to implement the BTS scheme.

Although the time-transformation function proposed by Tse and Dong (2014) can address the issue of intraday periodicity of transaction activity, it may not be

[3]As there are 6.5 hours of trades in a trading day for the NYSE, $t = 0, 1, \cdots, 23400$. $Q(t)$ is an increasing function of t, with $Q(0) = 0$ and $Q(23400) = 1$.

appropriate for BT sampling if volatility and transaction activity exhibit different intraday periodicity patterns. Figure 3.1 presents the intraday periodicity in volatility and trading activity of the stock JP Morgan (JPM) from January 2010 to April 2013. Figure 1A plots the means of the 1-min intraday realized volatilities over all trading days in the sample period, expressed in annualized standard deviation in percent, while Figure 1B plots the total number of transactions at each second from 09:30 to 16:00 over all trading days in the sample. We observe that the realized volatility at the beginning of the trading day is approximately two to three times larger than that near the end of the trading day. In contrast, the number of transactions at the end of the trading day is two to three times larger than that in the morning. This finding is quite regular across other stocks in our sample, which shows that intraday periodicity patterns in volatility and trading activity are quite different.[4] Thus, TTS returns in the morning will have larger volatility than those in the afternoon. In contrast, by construction BTS returns will have nearly constant volatility over the whole trading day.

3.2.2 Implementing the Business Time Sampling (BTS) Scheme

To implement the BTS scheme we propose to compute the time-transformation function $Q(t)$ using intraday integrated volatility estimated by a jump-robust method. The inverse function $Q^{-1}(t)$ is then used to obtain BTS transactions at a given target average sampling frequency. We outline our proposed method to obtain the BTS transactions as follows. Let X_t be the stock price process and $Y_t = \log X_t$ be the log-price process, sampled at calendar time points $t_1 < t_2 < \cdots < t_n$. We denote the full grid containing all observed points by $\mathscr{G} = \{t_1, t_2, \cdots, t_n\}$ and any arbitrary subgrid of $\mathscr{G}$ by $\mathscr{H}$. If $t_j \in \mathscr{H}$, then $t_{j,-}$ and $t_{j,+}$ denote the preceding and following elements of t_j in $\mathscr{H}$, respectively. We define $|\mathscr{H}|$ as the number of time increments in an arbitrary grid $|\mathscr{H}|$, so that $|\mathscr{H}| =$ (number of points in grid $\mathscr{H}$) $- 1$.

Let T denote the calendar-time length in seconds aggregated over all trading

[4]Results for other stocks can be found in the supplementary material.

days in the sample. Thus, for m trading days we have $T = 23400m$: Consider a sequence of estimated intraday integrated volatility $V_k, k = 1, \cdots, K$ for K consecutive time intervals over the period $(0, T]$, with the end point in each time interval represented by t_k, for $k = 1, \cdots, K$.[5] Denote the collection of points $t_k, k = 0, \cdots, K(t_0 = 0)$ by $\mathscr{H}_V$ and define $N_{t_0} = 0$ and $N_{t_k} = \sum_{i=1}^{k} V_i$ for $k = 1, \cdots, K$. The time-transformation function is calculated as $Q(t_k) = N_{t_k}/N_{t_K}$, for $t_k \in \mathscr{H}_V, k = 0, 1, \cdots, K$. $Q(t)$ at any calendar-time point t can then be computed using a cubic interpolation that preserves monotonicity in t.[6] The diurnally transformed time corresponding to calendar time t is denoted by $\tilde{t}$, with $\tilde{t} = TQ(t)$. Conversely, given a diurnally transformed time $\tilde{t}$, the corresponding calendar time is $t = Q^{-1}(\tilde{t}/T)$.

To sample a sequence of calendar-time points with BTS duration h, we take equally spaced diurnally transformed BT points $\tilde{t}_j, j = 0, \cdots, L$, with $\tilde{t}_j - \tilde{t}_{j-1} = h$ and $L = [T/h]$. Then $t_j = Q^{-1}(\tilde{t}_j/T),\ j = 0, \cdots, L$ are the required corresponding calendar-time points for the BTS scheme.[7] The return over each BT sampled interval will then have approximately equal volatility.

Figure 3.2 presents the time-transformation functions for stock JPM based on intraday trading activity and volatility. These functions are computed by merging the data over the complete sample period, resulting in representative one-day time-transformation functions.[8] Note that the two transformation functions exhibit different intraday patterns, with the compression of diurnally transformed time during market open more prominent for volatility than for trading activity.

[5] Here t_k are calendar-time points which need not to be regularly spaced. We outline the detailed steps in calculating V_k and t_k, for $k = 1, \cdots, K$, in the Appendix. In this paper we use the Tripower Realized Volatility (TRV) method of Barndorff-Nielsen *et al.* (2006) to calculate V_k.

[6] We use the Matlab command *pchip* in this paper.

[7] Denote the collection of t_j , for $j = 0, 1, \cdots, L$ by $\mathscr{H}_0$. Empirically, there may be no transaction at the selected time point t_j , which means $t_j \notin \mathscr{G}$. Instead of creating transactions using the previous-tick method, we select the nearest transaction after time t_j , i.e., at $t_j^* = \{\min\{t_i\} | t_i \in [t_{j,-}, t_j)$ for $t_{j,-}, t_j \in \mathscr{H}_0$ and $t_i \in \mathscr{G}\}, j = 1, \cdots, L$. The collection of t_j^* are the final BTS transaction calendar time points and we denote them by $\mathscr{H}_V$. If there is no transaction lying within the time interval selected, then no transaction t_j^* is selected from that time interval.

[8] We use the Matlab command *pchip* in this paper.

3.3 Testing the Semi-martingale Hypothesis using BTS Returns

As discussed above, BTS returns are iid normal for a semi-martingale price process even when there is leverage and/or feedback effect, whereas the CTS returns may not be iid normal even if the price process is a continuous local martingale. We now examine empirically the behavior of the BTS returns following the study of Andersen *et al.* (2010).

3.3.1 The Semi-martingale Hypothesis

Let $r_k = Y_{t_k} - Y_{t_{k-1}}$ be the jump-adjusted returns over the time interval (t_{k-1}, t_k).[9] If the log-price follows a jump-diffusion process with no leverage and volatility feedback, r_k standardized by the integrated volatility will follow a standard normal distribution, i.e.,

$$\frac{r_k}{\left(\int_{t_{k-1}}^{t_k} \sigma^2(\tau)d\tau\right)} \sim N(0,1), \quad k = 1, 2, \cdots, \tag{3.3.1}$$

where $\sigma(\cdot)$ is the instantaneous volatility function. The above result, however, will not hold if $\sigma(\cdot)$ exhibits correlation over time (feedback effect) or with the log-price innovation (leverage effect). On the other hand, if the jump-adjusted returns are sampled over business time (financial time) so that over each business-time interval $\tilde{t}_{k-1}, \tilde{t}_k$, we have

$$\tilde{t}_k = \underset{s>\tilde{t}_{k-1}}{inf} \left\{ \int_{\tilde{t}_{k-1}}^{s} \sigma^2(\tau)d\tau > \bar{\sigma}^2 \right\} \tag{3.3.2}$$

[9] We assume Y_t has been jump adjusted. The jump-adjustment procedure can be found in the Appendix.

for a given volatility threshold $\bar{\sigma}^2$, then the jump-adjusted returns $\tilde{r}_k$ over the business-time intervals $(\tilde{r}_{k-1}, \tilde{r}_k)$ satisfy

$$\frac{\tilde{r}_k}{\bar{\sigma}} \sim N(0,1), k = 1,2,\cdots. \tag{3.3.3}$$

To sample a sequence of BTS returns, Andersen *et al.* (2010) include intraday returns until the cumulative squared 5-min returns exceed a threshold, defined as the average daily or weekly realized volatility in calendar time (ABFN method hereafter). They report improved accuracy of the normal approximation under this sampling scheme.

3.3.2 Empirical Results of the Tests

To examine empirically the performance of our proposed BTS method versus the ABFN method, we use data of the top 40 market-capitalization stocks from the NYSE in 2010. We extract the tick-by-tick transaction data of these stocks from the TAQ database from January, 2010 to April, 2013. To clean the raw data, we follow the steps described in Tse and Dong (2014). Using the sequential jump-detection procedure of Andersen *et al.* (2010), we investigate cases of jumps over different sampling intervals and sampling schemes.[10] We test the normality assumption of the drift-corrected and jump-adjusted BTS returns. For each trading day, all BTS returns with jumps are deleted and the jump-adjusted daily or weekly BTS returns are then computed by summing remaining consecutive jump-adjusted BTS returns. We apply the Lilliefors test for normality to the jump-adjusted daily and weekly BTS returns and the returns obtained by the ABFN method. The results are reported in Table 1.

[10]Details of the selected stocks and results of the jump tests can be found in the supplementary material for which sampling frequencies of 1 min, 5 min and 10 min are used. When the sampling frequency is equal to 1 min, more than 12 stocks report jump proportions with values exceeding 10% under all sampling schemes. This suggests that sampling frequency which is too high (such as 1 min) may render misleading results when they are used for jump detection using the method of Andersen *et al.* (2010).

It can be seen that the BTS method improves the normality approximation of the standardized return distribution over the ABFN method. While the results for the weekly data are similar for the two methods, the BTS sampling scheme for the daily data restores normality for several stocks.

3.4 Estimation of Integrated Volatility

We now examine the use of the BTS scheme for estimating intraday volatility. We consider two methods of estimating daily volatility: Realized Volatility (RV) method and Autoregressive Conditional Duration-Integrated Conditional Volatility (ACD-ICV) method. The literature on RV estimation has grown tremendously since its inception. In this paper we select the Tripower Realized Volatility (TRV) estimate of Barndorff-Nielsen *et al.* (2006) for its robustness to price jumps. We compare the performance of the TRV estimates when returns are sampled by BTS, CTS and TTS schemes, with and without subsampling. With the same estimator used, the performance of the estimates is differentiated by the sampling method. We also consider the use of the ACD-ICV approach, with some modifications based on the BTS methodology.

3.4.1 Realized Volatility Estimation using BTS Returns

For a given subgrid H, the TRV estimate is computed as[11]

$$V_T = \xi_{\frac{2}{3}}^{-3} \left[\sum_{i=2}^{|\mathscr{H}|-1} |r_{i,-}|^{\frac{2}{3}} |r_i|^{\frac{2}{3}} |r_{i,+}|^{\frac{2}{3}} \right], \tag{3.4.1}$$

where $r_i = Y_{i,+} - Y_i$ for $Y_{i,+}, Y_i \in \mathscr{H}$ and $\xi_k = 2^{\frac{k}{2}} \Gamma((k+1)/2)/\Gamma(1/2)$ for $k > 0$, with $\Gamma(\cdot)$ denoting the gamma function.[12] For the CTS scheme, the previous-tick

[11] r_i here is used generically for CTS, TTS and BTS returns.

[12] There is an alternative finite-sample adjustment term $|\mathscr{H}|/(|\mathscr{H}|-2)$ for equation (3.4.1). This adjustment, however, uses the sample mean and may be downward biased when the intraday volatility exhibits strong asymmetry. Empirically, the intraday volatility at the beginning time of the trading day is much higher than that in other time periods, as shown in Figure 3.1. Thus, we adjust the

method is adopted when there is no transaction at the selected time point. For the TTS scheme, the number of subsampling grids S is selected to ensure that each subgrid has transactions at the target sampling frequency. To implement the subsampling method under the BTS scheme, we select BTS transactions with the sampling frequency being twice the average transaction duration. Subsampling is then implemented to obtain subgrids at the target sampling frequency.

3.4.2 The Modified ACD-ICV Method

Tse and Yang (2012) propose the ACD-ICV method to estimate daily and intraday volatility by modeling the price durations parametrically. Their simulation results show that the ACD-ICV method performs better than other methods (such as the Realized Kernel (RK) method of Barndorff-Nielsen *et al.* (2008)). The ACD-ICV method samples observations from the observed transaction data based on a pre-specified price threshold δ. Suppose $\mathscr{H}_{PE} = \{t_0, t_1, t_2, \cdots, t_N\}$ is the selected price events and the ith price duration is $x_i = t_i - t_{i-1}, i = 1, \cdots, N$. Let Φ_i denote the information set upon the price event at time t_i. Denote $\psi_{i+1} = E(x_{i+1}|\Phi_i)$ as the conditional expectation of the price duration and assume that the standardized durations $\varepsilon_i = x_i/\psi_i, i = 1, \cdots, N$ are iid positive random variables with mean of unity. Given the information Φ_i at time t_i, the conditional instantaneous return variance per unit time at time $t > t_i$, denoted by $\sigma^2(t|\Phi_i)$, is

$$\sigma^2(t|\Phi_i) = \frac{\delta^2}{\psi_{i+1}} \lambda \left(\frac{t - t_i}{\psi_{i+1}} \right), \tag{3.4.2}$$

where $\lambda(\cdot)$ is the hazard function of ε_i. Assuming ε_i to be iid standard exponential distributed, the integrated conditional variance (ICV) over time period $[t_{n_1}, t_{n_2+1}]$ is calculated as

$$ICV = \delta^2 \sum_{i=n_1}^{n_2} \left[\frac{t_{i+1} - t_i}{\psi_{i+1}} \right]. \tag{3.4.3}$$

finite-sample bias by adding the term $|r_{2,-}|^{\frac{2}{3}}|r_2|^{\frac{2}{3}}|r_{2,+}|^{\frac{2}{3}}$ in equation (3.4.1). Our simulation results confirm the better performance of this adjustment.

The conditional expectation of the price durations ψ_i can be estimated by various methods, such as the ACD method of Engle and Russell (1998) or the Augmented ACD method of Fernandes and Grammig (2006).

In this paper, we modify the ACD-ICV method in two ways. First, instead of modeling the durations of the price events obtained by the threshold δ, we model the BTS durations with the BTS returns sampled as in Section 2. Second, we replace δ^2 in equation (3.4.1) by an estimate of the mean volatility over each sampled BTS return. Suppose there are K BTS returns over m trading days, and the estimated integrated volatility over these m trading days is equal to V_m. Then each BTS return has an approximately constant integrated volatility of $V_D = V_m/K$. Instead of using δ^2 as an approximation of the integrated volatility of each price event as in Tse and Yang (2012), we use V_D to replace δ^2 in equation (3.4.1) to obtain a new ACD-ICV estimate.

Thus, for the ACD-ICV approach we consider three variations of estimates. We first estimate the daily integrated volatility using the ACD-ICV method as in Tse and Yang (2012) and denote this method by ME1. We then replace δ^2 by V_D, which is the integrated volatility estimated using TRV with subsampling at 3-min sampling frequency. We call this method ME2.[13] Finally, we sample data using the BTS scheme (not by price events) and repeat the computation as in ME2, which is called ME3. We compare the daily volatility estimates using the ACD-ICV methods against the RK method.[14]

[13]Note that for ME1 and ME2, the returns are sampled by price events and the ACD models are fitted to diurnally transformed durations using the time-transformation function based on the number of trades.

[14]The RK method is selected for comparison due to its superior performance among the RV estimators (see Barndorff-Nielsen *et al.* (2008)). To calculate the bandwidth of the RK method, we use the subsampling realized volatility estimator and 3-min TTS returns. For the ACD-ICV methods, all results in this paper are based on conditional duration models fitted using the power ACD (PACD) model (see Fernandes and Grammig (2006)).

3.5 Monte Carlo Study

We conduct a Monte Carlo (MC) study to examine the performances of different intraday volatility estimates. Our MC set-up draws upon other models in the literature.

3.5.1 Simulation Models

We consider five simulation models, which are summarized in Table 2. Model MD1 and MD2 are the Heston models (Aït-Sahalia and Mancini (2008) with some modifications) with high and low volatility, respectively. MD3 is the two-factor afine stochastic volatility model with an intraday U-shape pattern (Hasbrouck (1999) and Andersen *et al.* (2012)). MD4 is a deterministic volatility set-up (Tse and Yang (2012)). Finally, MD5 is MD1 with price jumps.

For all set-ups above, we set the initial price to 60 and the initial value of σ to 30%. We introduce sparsity of trade to the data by simulating exponentially distributed calendar-time transaction durations.[15] We first simulate transactions sec by sec and then generate exponentially distributed transaction durations with mean equal to 5 sec, 10 sec and 20 sec, respectively. For simplicity we only investigate iid market microstructure noise with constant noise-signal ratio (NSR). Based on the findings in Dong and Tse (2014), we consider cases with NSR = 0.005%, 0.01% and 0.02%. We introduce a 0.01 price rounding error in all simulations. The intraday duration periodicity is adjusted by the time-transformation method in Tse and Dong (2014) before we fit all price durations and BTS durations to the ACD model. Each model is simulated over 60 trading days, with the simulation repeated 1000 times.

[15] Sparsity occurs as empirically transactions are not observed sec by sec. Inactive stocks typically have more sparse transactions.

3.5.2 Simulation Results

We first report our results on the TRV estimates. For each model set-up sampling frequencies of 1, 2, 3, 5 and 10 min are considered. We compute TRV based on the CTS, TTS and BTS returns, with and without subsampling. All results can be found in the supplementary material. The BTS returns perform the best in reporting generally smaller root mean-squared error (RMSE), especially for methods with no subsampling. When the subsampling method is used, the RMSE decreases significantly. Generally the BTS scheme still performs the best and its advantage is especially significant for MD3 and MD4 when there is intraday volatility periodicity in the simulated price process.[16] Under all sampling schemes, the TRV estimates suffer from the market microstructure noise problems when the NSR is large, resulting in high mean error (ME) at high sampling frequency. When sparsity and NSR are both low, high sampling frequency at 1 min interval produces the lowest RMSE.

Overall, the BTS returns outperform the CTS and TTS returns in estimating the integrated volatility. To save space we present only the results for the BTS scheme with subsampling in Table 3. Table 4 summarizes the average difference in RMSE of the CTS and TTS schemes versus the BTS scheme over all model and parameter set-ups, with subsampling, in both absolute and relative terms. It can be seen that better intraday volatility estimates can be obtained by adopting the BTS scheme when the TRV method is used.

We now turn to the results of the ACD-ICV method. Table 5 and Table 6 report the ME and RMSE of the RK and the ACD-ICV estimates for MD1 and MD5, respectively. Results for other models can be found in the supplementary material.[17] The RK method performs quite well for the unbiasedness property, reporting small

[16] When there is intraday volatility periodicity the BTS returns resemble more closely to normal distribution than the CTS and TTS returns.

[17] For the ACD-ICV method, ME1 and ME2 sometimes report rather large estimation errors for the first two days of the simulation, which is probably due to the starting values of the estimated conditional expected durations of the fitted PACD model of the price durations. Hence, we drop the first two days for these two methods when calculating ME and RMSE.

ME for all models except MD5 (model with price jump). ME1 reports quite big absolute ME and RMSE values among all the sampling frequencies considered and it often performs worse than the RK method except for MD5.[18] In contrast, the modified ACD-ICV methods, ME2 and ME3, report quite small ME and RMSE. ME3 consistently reports smaller RMSE than ME2 except for few cases in MD3 and MD5 at the 3-min and 5-min sampling frequencies. Moreover, ME2 varies more across different sampling frequencies. This demonstrates the superiority of the BTS durations over the price durations when they are fitted to the ACD model to estimate the integrated volatility. The better performance of ME3 over ME2 is mainly due to the fact that the BTS scheme performs better in yielding returns with constant volatility compared against the ACD-ICV method using the price events.

Our modified ACD-ICV method ME3 consistently produces lower RMSE over all models and parameter set-ups than the RK estimates. Its performance is robust over a wide range of sampling frequencies of up to 15 min. It also outperforms the TRV estimates with subsampling, except possibly for MD5, in which case their performances are comparable.[19]

3.6 Conclusion

We propose an easy-to-use time-transformation method to implement the BTS scheme at a prespecified average sampling frequency. Using 40 stocks from the NYSE, we perform normality test to the jump-adjusted daily and weekly BTS returns. Our results show that stock prices can be considered discrete observations from a continuous-time jump-diffusion process. The BTS scheme performs better

[18]This is in contrast to the findings in Tse and Yang (2012), which shows the superiority of the ACD-ICV method over the RK method via simulation using sec-by-sec transactions (sparsity of 1 sec). The poor performance of ME1 is mainly due to the transaction sparsity, since using σ^2 as the proxy for integrated volatility over one price event becomes unreliable when transactions are sparse. Supporting evidence is provided in our simulation study that the RMSE of ME1 increases when observed transactions are more sparse.

[19]MD5 is a model with price jumps, and the TRV method is constructed to be robust to price jumps.

than the CTS and TTS schemes in yielding iid Gaussian returns, and it also performs better than the CTS and TTS schemes in estimating daily integrated volatility using the TRV method, with and without subsampling. We also show the superiority of the BTS durations over the price durations in estimating the daily integrated volatility using the ACD-ICV method. Our modified ACD-ICV estimate, ME3, which models the high-frequency BTS durations using the ACD model, performs the best in reporting smaller RMSE values.

Finally, we note that there are other possible applications of the BTS scheme. For example, we can estimate the conditional instantaneous volatility or intraday integrated volatility (such as over 30-min intervals) by modeling the high-frequency BTS durations using the ACD model. Moreover, using a method similar to that in obtaining BTS returns, we can obtain transactions with approximately equal quarticity increments. The conditional instantaneous quarticity or integrated quarticity can be estimated further by modeling clustering of the corresponding durations using the ACD model.

Table 3.1: Test of normality hypothesis for no-jump returns of 40 NYSE stocks

	ABFN		BTS	
	5%	1%	5%	1%
Daily	12	5	6	4
Weekly	3	1	4	0

Notes: The figures are the numbers of stocks (out of 40) for which the normality hypothesis at the 5% and 1% levels of significance are rejected based on the Lilliefors test implemented for daily and weekly jump-adjusted returns, 2010-2013. The returns are constructed over 5-min sampling frequency using the method of Andersen et al. (2010) (ABFN) and the method proposed in this paper (BTS). The total number of trading days of each stock ranges from 830 to 853.

Table 3.2: Summary of simulation models

Model	Code	Description of Model	Description of Model Parameters
Heston Model (high volatility)	MD1	$d\log X(t) = \left(\mu - \frac{\sigma^2(t)}{2}\right) dt + \sigma(t) dW_1(t)$, $d\sigma^2(t) = \kappa(\alpha - \sigma^2(t))dt + \gamma\sigma(t)dW_2(t)$.	$\mu = 0.05$, $\kappa = 5$, $\alpha = 0.25$, $\gamma = 0.5$ and $\mathrm{corr}(dW_1(t), dW_2(t)) = -0.5$.
Heston Model (low volatility)	MD2	Same as MD1.	$\alpha = 0.04$, all remaining parameters same as MD1.
Two-factor affine stochastic volatility model with U-shape intraday volatility pattern	MD3	$d\log X(t) = \sigma_u(t)\sigma_{sv}(t)dW_1(t)$, $\sigma_{sv}^2(t) = \sigma_1^2(t) + \sigma_2^2(t)$, $d\sigma_1^2(t) = \kappa_1[\theta_1 - \sigma_1^2(t)]dt + \eta_1\sigma_1(t)dW_{21}(t)$, $d\sigma_2^2(t) = \kappa_2[\theta_2 - \sigma_2^2(t)]dt + \eta_2\sigma_2(t)dW_{22}(t)$, $\sigma_u(t) = C + Ae^{-at} + Be^{-b(1-t)}$, $t \in [0,1]$.	$\kappa_1 = 0.6$, $\kappa_2 = 0.1$, $\theta_1 = 0.09$, $\theta_2 = 0.04$, $\eta_1 = 0.2$, $\eta_2 = 0.1$, $\rho_1 = \mathrm{Corr}(dW_1(t), dW_{21}(t)) = 0.9$, and $\rho_2 = \mathrm{Corr}(dW_1(t), dW_{22}(t)) = -0.4$. $A = 0.75$, $B = 0.25$, $C = 0.88929198$, $a = 10$ and $b = 10$.
Deterministic volatility model with U-shape intraday volatility pattern	MD4	$d\log X(u) = \sigma(u)dW(u)$, $\sigma(u) = \sigma(t,\tau) = \sigma_1(t)\sigma_2(\tau)$, where t is the day of trade and τ is the intraday time. $\sigma_1(t)$ represents the average volatility of day t and $\sigma_2(\tau)$ captures intraday variations at time τ of each trading day.	$\sigma_1(t) = 20\%$ for $t = 1$ with $\sigma_1(t)$ increasing linearly in t over 20 days to reach 30%. It then remains level for the next 20 days and decreases linearly in t to 20% over 20 days. $\sigma_2(\tau)$ is computed as in Tse and Yang (2012) using the IBM tick-by-tick transaction data in 2012.
MD1 with price jumps	MD5	$d\log X(t) = \left(\mu - \frac{\sigma^2(t)}{2}\right) dt + \sigma(t) dW_1(t) + J(t)dP(t)$, $d\sigma^2(t) = \kappa(\alpha - \sigma^2(t))dt + \gamma\sigma(t)dW_2(t)$.	$P(t)$ is a Poisson process with constant mean such that on average one price jump happens every two simulated trading days. $J(t)$ refers to the size of the corresponding discrete jumps in the log-price process with $J(t) \sim \mathrm{N}(0.02, 0.004)$, with remaining set-up the same as MD1.

Table 3.3: ME and RMSE of daily volatility using the TRV method with subsampling under the BTS scheme

Sparsity	NSR	Model	ME					RMSE				
			1-min	2-min	3-min	5-min	10-min	1-min	2-min	3-min	5-min	10-min
5-sec	0.005%	MD1	-0.4286	0.0470	0.0031	0.0004	-0.1777	1.3869	1.7228	2.1610	2.8165	4.0975
		MD2	-0.4321	-0.0500	-0.0513	-0.0338	-0.1309	0.9964	1.1706	1.4665	1.9242	2.7797
		MD3	-0.4435	0.0200	0.0123	-0.0019	-0.1916	1.2425	1.5657	1.9431	2.5511	3.6823
		MD4	-0.2403	-0.0913	0.0126	0.0055	-0.2069	1.0574	1.3723	1.6849	2.2278	3.2164
		MD5	-0.0687	0.6282	0.7732	1.0940	1.6463	1.4163	2.0196	2.5587	3.4485	5.2161
	0.01%	MD1	0.3382	0.4486	0.2625	0.1532	-0.1142	1.3520	1.7646	2.1562	2.8077	4.0850
		MD2	0.0786	0.2172	0.1192	0.0675	-0.0884	0.9049	1.1831	1.4608	1.9152	2.7695
		MD3	0.3291	0.4002	0.2555	0.1439	-0.1301	1.2126	1.6016	1.9521	2.5460	3.6734
		MD4	0.5904	0.3128	0.2734	0.1532	-0.1474	1.1809	1.4043	1.7059	2.2373	3.2230
		MD5	0.7280	1.0331	1.0365	1.2492	1.7128	1.5989	2.1671	2.6391	3.4936	5.2306
	0.02%	MD1	1.8463	1.2523	0.7789	0.4499	0.0126	2.2762	2.1053	2.2506	2.8148	4.0590
		MD2	1.0802	0.7479	0.4603	0.2651	-0.0059	1.4503	1.3864	1.5109	1.9145	2.7529
		MD3	1.7862	1.1644	0.7397	0.4265	-0.0167	2.1440	1.9166	2.0542	2.5558	3.6603
		MD4	2.2616	1.1036	0.7889	0.4437	-0.0291	2.4966	1.7643	1.8609	2.2798	3.2357
		MD5	2.2713	1.8503	1.5649	1.5559	1.8424	2.7080	2.6518	2.8753	3.5982	5.2587
10-sec	0.005%	MD1	-0.9375	-0.0435	-0.0307	-0.0333	-0.1880	1.6335	1.8357	2.1600	2.8457	4.1021
		MD2	-0.8123	-0.1034	-0.0778	-0.0560	-0.1417	1.2304	1.2500	1.4735	1.9394	2.7833
		MD3	-0.8436	-0.0958	-0.0423	-0.0340	-0.2088	1.4854	1.6584	1.9742	2.5751	3.7013
		MD4	-0.5797	-0.2436	-0.0830	-0.0584	-0.2432	1.2715	1.4404	1.7121	2.2547	3.2308
		MD5	-0.5986	0.5429	0.7290	1.0637	1.6334	1.5659	2.0991	2.5573	3.4711	5.2171
	0.01%	MD1	0.0713	0.3397	0.2356	0.1224	-0.1251	1.3501	1.8561	2.1611	2.8355	4.0856
		MD2	-0.1345	0.1520	0.1010	0.0460	-0.0992	0.9467	1.2544	1.4661	1.9323	2.7757
		MD3	0.1028	0.2815	0.2097	0.1090	-0.1529	1.2390	1.6800	1.9774	2.5686	3.6909
		MD4	0.3144	0.1905	0.1929	0.0949	-0.1830	1.1801	1.4321	1.7293	2.2613	3.2356
		MD5	0.4190	0.9362	0.9980	1.2244	1.6968	1.5275	2.2311	2.6363	3.5126	5.2337
	0.02%	MD1	2.0479	1.1137	0.7770	0.4235	-0.0051	2.5005	2.1285	2.2661	2.8358	4.0650
		MD2	1.1815	0.6675	0.4586	0.2426	-0.0175	1.5885	1.4198	1.5277	1.9301	2.7608
		MD3	1.9622	1.0222	0.7158	0.3933	-0.0340	2.3465	1.9448	2.0766	2.5855	3.6804
		MD4	2.1082	1.0387	0.7207	0.3933	-0.0644	2.4205	1.7728	1.8823	2.2997	3.2520
		MD5	2.4111	1.7259	1.5437	1.5352	1.8277	2.8718	2.6669	2.8772	3.6225	5.2623
20-sec	0.005%	MD1	-1.9043	-0.4650	-0.1715	-0.1586	-0.3052	2.5154	1.9081	2.3632	2.8884	4.1176
		MD2	-1.4513	-0.3972	-0.1718	-0.1551	-0.2189	1.8465	1.3232	1.6142	1.9728	2.8010
		MD3	-1.7349	-0.4945	-0.1976	-0.1617	-0.3245	2.2880	1.7421	2.0996	2.6517	3.7326
		MD4	-1.3673	-0.5413	-0.2730	-0.2261	-0.3586	1.9372	1.6463	1.8101	2.3044	3.2759
		MD5	-1.5684	0.0539	0.6149	0.9467	1.5263	2.3452	2.0185	2.7000	3.4925	5.2178
	0.01%	MD1	-0.7727	0.0595	0.0795	-0.0012	-0.2474	1.8388	1.8658	2.3619	2.8691	4.1079
		MD2	-0.6930	-0.0470	-0.0025	-0.0515	-0.1798	1.3436	1.2747	1.6111	1.9649	2.7898
		MD3	-0.6715	0.0110	0.0681	-0.0090	-0.2678	1.6684	1.6857	2.1131	2.6393	3.7245
		MD4	-0.3601	-0.0577	0.0292	-0.0623	-0.3056	1.4354	1.5473	1.7854	2.3140	3.2870
		MD5	-0.4358	0.5872	0.8854	1.1072	1.5868	1.8245	2.1169	2.7876	3.5323	5.2285
	0.02%	MD1	1.4062	1.0834	0.5790	0.3089	-0.1229	2.2543	2.1942	2.4346	2.8722	4.0765
		MD2	0.7467	0.6389	0.3302	0.1612	-0.0953	1.4646	1.4674	1.6470	1.9576	2.7740
		MD3	1.3598	0.9923	0.5629	0.2802	-0.1573	2.1106	1.9812	2.2102	2.6445	3.7064
		MD4	1.6100	0.9172	0.6508	0.2535	-0.1811	2.1813	1.8069	1.9153	2.3599	3.2984
		MD5	1.7450	1.6299	1.4063	1.4237	1.7176	2.5623	2.6474	3.0191	3.6372	5.2592

Notes: ME and RMSE are of the annualized standard deviation in percentage. The average true daily integrated volatility is around 40% for MD1, 27% for MD2, 36% for MD3, 28% for MD4 and 40% for MD5. MD1 and MD2 are the Heston model at different volatility level. MD3 is the two-factor stochastic volatility model with intraday volatility periodicity. MD4 is the deterministic volatility model with intraday volatility periodicity and MD5 is the Heston model (MD1) with price jumps. The first column indicates the average duration of the observed simulated transactions. The sampling frequency of the BTS scheme equals to twice the average transaction duration.

Table 3.4: Comparison of RMSE of TRV estimate under different sampling schemes with subsampling

	1-min	2-min	3-min	5-min	10-min
Avg. RMSE (CTS-BTS)	0.3043	0.1666	0.1891	0.2215	0.1926
Avg. RMSE (CTS-BTS)/BTS (%)	17.8262	9.8515	9.6493	8.8626	5.7456
Avg. RMSE (TTS-BTS)	-0.0237	0.1261	0.1721	0.2240	0.2072
Avg. RMSE (TTS-BTS)/BTS (%)	-0.1011	7.9977	9.0710	9.1005	6.2101

Notes: RMSE is the root-mean squared error of daily volatility estimation annualized standard deviation in percentage. The sampling frequency of the BTS scheme equals to twice the average transaction duration.

Table 3.5: ME and RMSE of daily volatility estimates of the RK and ACD-ICV methods for Model MD1

		ME							RMSE						
Sparsity	NSR	RK	ACD-ICV	Avg. sampling frequency (ACD-ICV)					RK	ACD-ICV	Avg. sampling frequency (ACD-ICV)				
				1-min	3-min	5-min	10-min	15-min			1-min	3-min	5-min	10-min	15-min
5-sec	0.005%	-0.1087	ME1	-5.8450	-3.4421	-2.6778	-1.8270	-1.3546	2.2978	ME1	6.2142	3.8881	3.3149	2.9738	2.9941
			ME2	-0.2043	-0.2366	-0.2205	-0.0928	0.0734		ME2	1.7527	1.7500	1.9673	2.3810	2.6985
			ME3	-0.1120	0.0540	0.2446	0.4362	0.7306		ME3	1.2772	1.4001	1.7261	1.4380	1.5706
	0.01%	-0.0876	ME1	-3.4503	-2.0139	-1.5554	-1.0491	-0.7357	2.2998	ME1	3.8811	2.6241	2.4551	2.5735	2.7766
			ME2	0.0437	0.0089	0.0301	0.1633	0.3410		ME2	1.6266	1.6906	1.9326	2.3896	2.7213
			ME3	0.1489	0.3076	0.4972	0.6950	0.9897		ME3	1.2367	1.3710	1.7078	1.4948	1.6735
	0.02%	-0.0461	ME1	1.4407	0.6924	0.4711	0.3438	0.3824	2.3040	ME1	2.0920	1.8468	2.0100	2.4290	2.7559
			ME2	0.5181	0.5045	0.5353	0.6804	0.8449		ME2	1.5778	1.7868	2.0340	2.5118	2.8656
			ME3	0.6677	0.8253	1.0079	1.2071	1.5051		ME3	1.3320	1.4939	1.8291	1.7399	1.9751
10-sec	0.005%	-0.1416	ME1	-10.9555	-6.0669	-4.6302	-3.1722	-2.4450	2.6341	ME1	11.3935	6.4347	5.0926	3.9820	3.6161
			ME2	-0.3400	-0.3899	-0.3720	-0.2555	-0.0909		ME2	2.2864	1.8765	2.0363	2.4242	2.6968
			ME3	-0.2235	-0.1976	0.0212	0.2764	0.5716		ME3	1.4774	1.4302	1.5870	1.4764	1.5774
	0.01%	-0.1178	ME1	-8.5745	-4.8751	-3.7310	-2.5558	-1.9404	2.6363	ME1	8.9899	5.2580	4.2561	3.4911	3.2988
			ME2	-0.0858	-0.1337	-0.1165	0.0015	0.1719		ME2	2.0281	1.7825	1.9930	2.4033	2.7120
			ME3	0.0409	0.0713	0.2829	0.5387	0.8356		ME3	1.4179	1.3836	1.5621	1.5129	1.6586
	0.02%	-0.0710	ME1	-4.8863	-2.6923	-2.0864	-1.4154	-1.0327	2.6399	ME1	5.3222	3.2140	2.8523	2.7546	2.8704
			ME2	0.4241	0.3767	0.3976	0.5304	0.7014		ME2	1.8341	1.7717	2.0131	2.4689	2.8195
			ME3	0.5665	0.5917	0.8060	1.0634	1.3599		ME3	1.4587	1.4454	1.6547	1.7292	1.9392
20-sec	0.005%	-0.1987	ME1	-17.7796	-9.5646	-7.3266	-5.0140	-3.9444	3.0600	ME1	18.3065	9.9612	7.7382	5.6130	4.7999
			ME2	-0.6770	-0.7315	-0.7231	-0.6051	-0.4518		ME2	3.1408	2.2301	2.2493	2.5140	2.7698
			ME3	-0.4760	-0.5490	-0.3795	-0.0762	0.2182		ME3	2.2789	1.6071	1.7431	1.5932	1.6135
	0.01%	-0.1723	ME1	-16.2877	-8.6764	-6.6429	-4.5336	-3.5434	3.0621	ME1	16.8362	9.0603	7.0596	5.1679	4.4519
			ME2	-0.4067	-0.4638	-0.4586	-0.3355	-0.1811		ME2	2.9539	2.0830	2.1464	2.4481	2.7132
			ME3	-0.2070	-0.2778	-0.1000	0.1926	0.4882		ME3	2.2062	1.5119	1.7034	1.5809	1.6494
	0.02%	-0.1200	ME1	-12.8662	-7.0624	-5.3456	-3.6371	-2.8233	3.0670	ME1	13.3676	7.4466	5.7892	4.3706	3.9009
			ME2	0.1191	0.0629	0.0728	0.2021	0.3562		ME2	2.6226	1.9304	2.0461	2.4303	2.7561
			ME3	0.3287	0.2614	0.4493	0.7287	1.0260		ME3	2.1632	1.4769	1.7526	1.7006	1.8539

Notes: ME and RMSE are of the annualized standard deviation in percentage. MD1 is the Heston model and the average true daily integrated volatility is around 40%. ME1 is the ACD-ICV method in Tse and Yang (2012). ME2 is ME1 with δ^2 replaced by V_D, the integrated volatility estimated using TRV with subsampling at 3-min sampling frequency. ME3 is ME2 with sampled durations computed from BTS returns. All ACD models are fitted to diurnally transformed durations using the time-transformation function based on the number of trades as in Tse and Dong (2014).

Table 3.6: ME and RMSE of daily volatility estimates of the RK and ACD-ICV methods for Model MD5

Sparsity	NSR	ME							RMSE						
		RK	ACD-ICV	Avg. sampling frequency (ACD-ICV)					RK	ACD-ICV	Avg. sampling frequency (ACD-ICV)				
				1-min	3-min	5-min	10-min	15-min			1-min	3-min	5-min	10-min	15-min
5-sec	0.005%	5.4227	ME1	-5.8297	-3.3882	-2.5730	-1.6045	-1.0188	9.9602	ME1	6.2026	3.8409	3.2360	2.8656	2.8789
			ME2	0.8058	0.7685	0.7892	0.9164	1.0902		ME2	1.9098	1.9199	2.1430	2.6091	2.9672
			ME3	0.8846	1.0464	1.2556	1.4577	1.7653		ME3	1.6636	1.8671	2.1671	2.0848	2.2923
	0.01%	5.4418	ME1	-3.4340	-1.9407	-1.4617	-0.8313	-0.3769	9.9686	ME1	3.8603	2.5721	2.4208	2.5277	2.7279
			ME2	1.0512	1.0187	1.0453	1.1758	1.3516		ME2	1.9215	2.0070	2.2572	2.7266	3.0797
			ME3	1.1471	1.3082	1.5160	1.7153	2.0254		ME3	1.7863	1.9921	2.2839	2.2406	2.4745
	0.02%	5.4794	ME1	1.4728	0.7713	0.5849	0.5818	0.7257	9.9860	ME1	2.0990	1.8775	2.0612	2.5042	2.8495
			ME2	1.5310	1.5181	1.5467	1.6891	1.8715		ME2	2.1539	2.3315	2.5607	3.0135	3.3723
			ME3	1.6721	1.8222	2.0263	2.2336	2.5459		ME3	2.1213	2.3145	2.5816	2.6222	2.8867
10-sec	0.005%	5.3907	ME1	-10.9449	-6.0276	-4.5318	-2.9491	-2.1042	10.0436	ME1	11.3795	6.4015	5.0126	3.8136	3.4271
			ME2	0.6633	0.6106	0.6242	0.7470	0.9182		ME2	2.3104	1.9461	2.1377	2.5742	2.9195
			ME3	0.7705	0.7948	0.9985	1.2914	1.5981		ME3	1.6828	1.8435	2.0209	2.0229	2.2163
	0.01%	5.4121	ME1	-8.6446	-4.8146	-3.6363	-2.3242	-1.5842	10.0534	ME1	9.0607	5.2047	4.1744	3.3347	3.1353
			ME2	0.9153	0.8660	0.8891	1.0125	1.1800		ME2	2.1941	1.9921	2.2058	2.6559	3.0205
			ME3	1.0358	1.0604	1.2689	1.5548	1.8631		ME3	1.7876	1.9431	2.1333	2.1752	2.3929
	0.02%	5.4548	ME1	-4.8774	-2.6232	-1.9874	-1.1786	-0.6562	10.0727	ME1	5.3053	3.1592	2.7833	2.6525	2.7837
			ME2	1.4283	1.3848	1.4085	1.5407	1.7137		ME2	2.2591	2.2408	2.4554	2.9128	3.2801
			ME3	1.5657	1.5953	1.7983	2.0797	2.3929		ME3	2.0947	2.2393	2.4394	2.5464	2.8021
20-sec	0.005%	5.3322	ME1	-17.8068	-9.5419	-7.2562	-4.8115	-3.5992	10.1532	ME1	18.3394	9.9398	7.6740	5.4435	4.5299
			ME2	0.3397	0.2821	0.2882	0.4019	0.5596		ME2	3.0755	2.1117	2.1666	2.5271	2.8374
			ME3	0.5340	0.4639	0.6157	0.9473	1.2592		ME3	2.2664	1.7445	1.9529	1.9189	2.0788
	0.01%	5.3562	ME1	-16.4046	-8.6377	-6.5612	-4.3376	-3.2101	10.1639	ME1	16.9502	9.0284	6.9857	5.0049	4.2219
			ME2	0.6035	0.5473	0.5600	0.6682	0.8417		ME2	2.9903	2.1008	2.1818	2.5643	2.9096
			ME3	0.8043	0.7385	0.8828	1.2168	1.5292		ME3	2.3152	1.8216	2.0152	2.0450	2.2400
	0.02%	5.4036	ME1	-12.8619	-7.0131	-5.2494	-3.4318	-2.4897	10.1854	ME1	13.3635	7.4026	5.6999	4.2129	3.6884
			ME2	1.1378	1.0778	1.0905	1.2102	1.3818		ME2	2.8324	2.2176	2.3262	2.7620	3.1219
			ME3	1.3435	1.2746	1.4297	1.7565	2.0717		ME3	2.5068	2.0651	2.2905	2.3864	2.6243

Notes: ME and RMSE are of the annualized standard deviation in percentage. MD5 is the Heston model with price jumps and the average true daily integrated volatility is around 40%. ME1 is the ACD-ICV method in Tse and Yang (2012). ME2 is ME1 with δ^2 replaced by V_D, the integrated volatility estimated using TRV with subsampling at 3-min sampling frequency. ME3 is ME2 with sampled durations computed from BTS returns. All ACD models are fitted to diurnally transformed durations using the time-transformation function based on the number of trades as in Tse and Dong (2014).

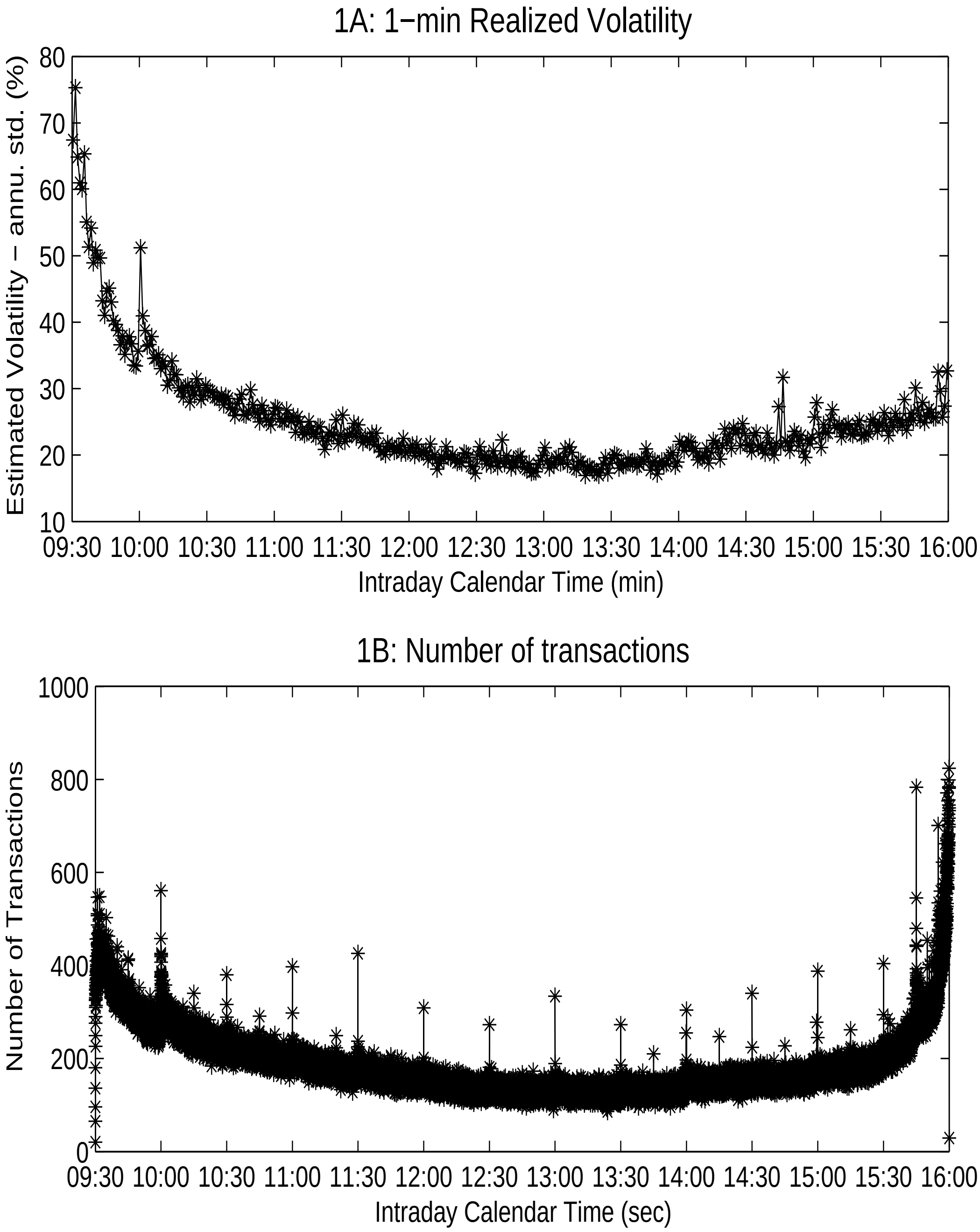

Figure 3.1: Intraday volatility and trading activity of JPM, 2010/01-2013/04.

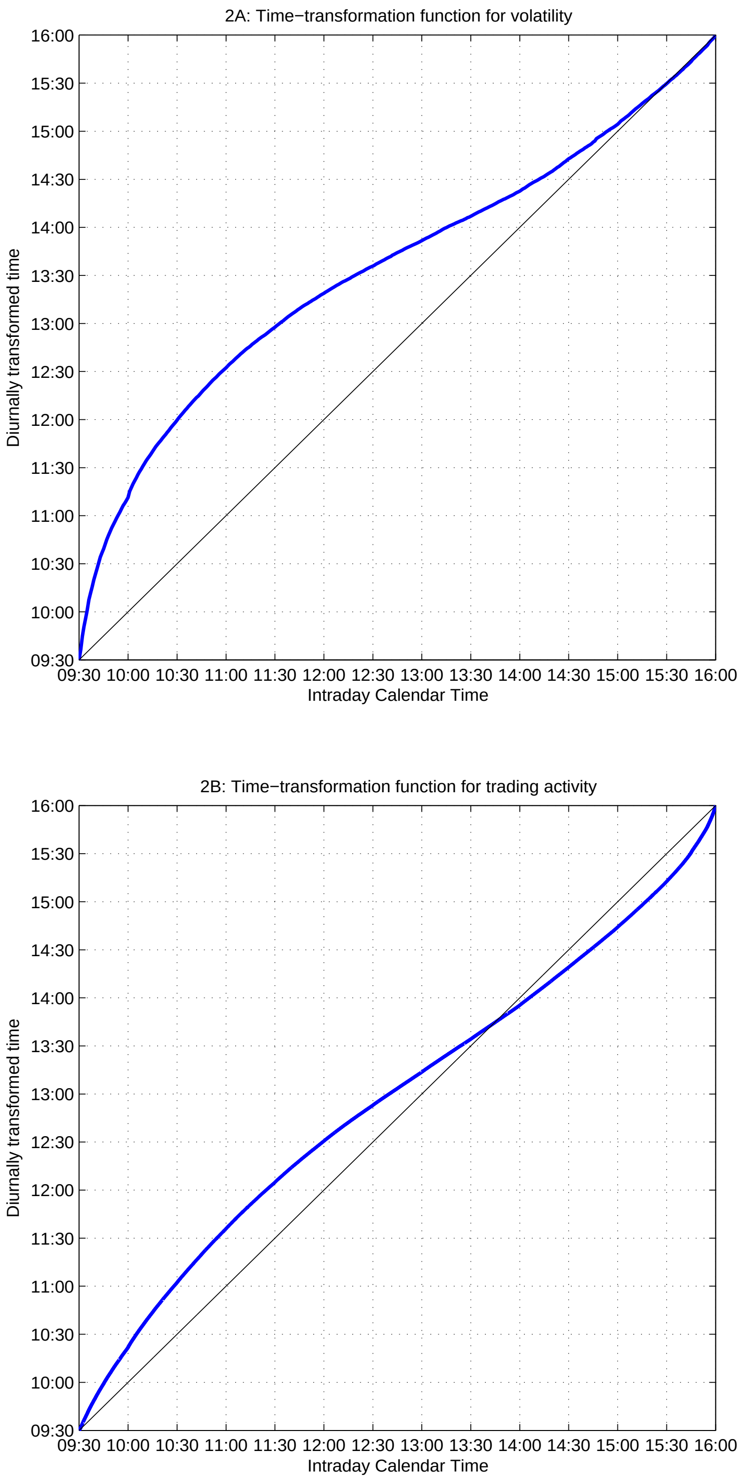

Figure 3.2: Time-transformation functions for JPM.

Chapter 4 A Study on Market Microstructure Noise

4.1 Introduction

The presence of market microstructure noise is a well-known phenomenon in financial markets and causes problems in high-frequency financial data analysis. Analysis of market microstructure noise properties is important for both high-frequency traders and market makers. Hasbrouck (1993) defines pricing error as the difference between the logarithmic observed transaction prices and the latent efficient prices. He points out that pricing error is a good measure of market quality and can be used to evaluate broker's performance or determine markets or regulatory structures under which transaction costs are minimized.

Various papers investigate the influence of microstructure noise in high-frequency financial data on volatility estimates of asset returns. For example, Andersen *et al.* (2000) demonstrate the bias introduced by the microstructure noise in the average realized volatility (RV) using the volatility signature plot. Andersen *et al.* (2001a, 2001b, 2003) and Barndorff-Nielsen and Shephard (2002a, 2002b) suggest using 5-min to 20-min sampling frequencies to alleviate the microstructure noise effect in estimating asset return volatility. Aït-Sahalia *et al.* (2005) suggest sampling a continuous-time process as often as possible. Andersen *et al.* (2011) and Ghysels and Sinko (2011) investigate the influence of the microstructure noise on volatility forecasting. Diebold and Strasser (2013) investigate the correlation structure of microstructure noise between latent asset returns and explore the application of the microstructural information to volatility estimation.

Hansen and Lunde (2006) investigate the empirical properties of microstructure noise using transaction prices and bid-ask quotes of 30 Dow Jones Industrial Average (DJIA) stocks. They find the noise to be correlated with the efficient price. They also use the noise-to-signal ratio (NSR) to characterize the magnitude of the noise and find NSR to be below 0.1% for DJIA transaction data in 2000.[1] Zhou (1996), Zhang *et al.* (2005), Hansen and Lunde (2006), Oomen (2006a) and Bandi and Russell (2008), among others, use NSR to characterize the mean squared-error (MSE) properties of realized variance estimates, which is then used to obtain the optimal sampling frequencies. Awartani *et al.* (2009) test the impact of the microstructure noise on volatility estimates by comparing realized volatilities computed at different sampling frequencies. They also test the hypothesis that the noise variance is independent of the chosen sampling frequencies by comparing estimated microstructure-noise variance at different sampling frequencies. Ubukata and Oya (2009) assume the noise covariance depends on the calendar time and examine the degree of dependence in microstructure noise by testing the cross- and auto-covariance estimates using the multiplication of two adjacent returns. Jacod *et al.* (2013) assume the noise covariance to be dependent on the number of transactions and estimate the arbitrary order of moments and joint moments of the microstructure noise using the pre-averaging method. All four papers aforementioned assume the influence of the cross correlation between the noise and asset returns at all sampling frequencies to be negligible when estimating the noise variance or auto-covariance.

In this paper we show that sampling at either very high or very low frequencies renders unreliable noise-variance estimates. First, noise is correlated with latent stock return at high sampling frequency, and its influence is not negligible. Second, we find the noise-variance estimates to have large estimation error for samples at low sampling frequencies. To alleviate these problems, we propose a

[1]NSR is defined as the ratio of microstructure-noise variance to daily stock return realized variance.

microstructure-noise variance estimator using subsampling realized variance at two different sampling frequencies. For this estimate, we can sample returns at two suitable frequencies, without necessarily using tick-by-tick returns or returns at 20-min or 30-min frequencies, as commonly used in the literature. In doing this, we can mitigate the estimation error introduced by too high or too low sampling frequencies. In practice, we suggest sampling transactions at moderately high frequencies, such as 1 min or 2 min.

We further propose an estimate using subsampling realized variance at multiple time scales. At each time point we calculate the subsampling realized variance at sampling frequencies centering around it. We then estimate the noise variance using the difference of weighted average subsampling realized variances. An advantage of this estimate is that we can utilize information at multiple time scales, rather than only two time scales. We show that this estimate has smaller variance than the one using two time scales.

We compare different noise-variance estimates using Monte Carlo simulation study and find that our proposed multi-scale estimates have smaller mean error (ME) and root mean-squared error (RMSE). Using this estiamte, we find that the NSR is around 0.005% for most of the selected New York Stock Exchange (NYSE) transaction data from 2010 to 2013. This is significantly smaller than the estimate of 0.1% reported by Hansen and Lunde (2006) for the DJIA transaction data in 2000. Differences in the calculated noise variances stem from different sources. First, we use different noise-variance estimates from those in Hansen and Lunde (2006). We show that our methods tend to be more reliable, as demonstrated by our Monte Carlo study. Second, the time span between the two data set is large. O'Hara (2015) point out that due to technological developments and regulatory policy changes, spreads are smaller now because a lot of high frequency trading is based on intermarket arbitrage.[2]

[2] Such as the changes in the U.S. Regulation Alternative Trading Systems in 2000 and the Regulation National Market System in 2007.

The rest of the paper is organized as follows. Section 2 outlines some assumptions of the latent price process and the microstructure noise. We also describe five noise-variance estimates in this section. In Section 3 we characterize the microstructure noise properties of the selected NYSE stocks. We conduct a Monte Carlo study in Section 4 to compare different noise-variance estimates. We report estimated NSR of selected NYSE stocks in Section 5 and draw conclusions in Section 6. The asymptotic distributions of the noise-variance estimates are derived in the Appendix.

4.2 Microstructure Noise-Variance Estimates

4.2.1 Realized Variance Estimates

Let $\{X_t\}$ denote a latent efficient log-price process in continuous time and $\{Y_t\}$ denote the observed log-price process. Following Hasbrouck (1993), the noise process is $\{\varepsilon_t\}$, with $\varepsilon_t = Y_t - X_t$. We define the full grid containing all sampled time points (in sec) by $\mathscr{G} = \{t_0, t_1, t_2, \cdots, t_n\}$ and any subgrid of $\mathscr{G}$ by $\mathscr{H}$.[3] If $t_i \in \mathscr{H}$, let t_{i-} and t_{i+} denote the preceding and following elements of t_i in $\mathscr{H}$, respectively. Let $|\mathscr{H}|$ be the number of time increments in subgrid $\mathscr{H}$, so that $|\mathscr{H}| =$ (number of points in subgrid $\mathscr{H}$) $- 1$. The realized volatility of a generic process Z ($Z = Y$ or X) on $\mathscr{H}$ is given by

$$[Z,Z]_{\mathscr{H}} = \sum_{t_i,\ t_{i-} \in \mathscr{H}} (Z_{t_i} - Z_{t_{i-}})^2. \tag{4.2.1}$$

Empirically we only observe Y_t, while the presence of the microstructure noise ε_t influences the stock return volatility estimates. To alleviate the noise effect, trans-

[3] In practical applications, transactions can be sampled under different sampling schemes, such as Tick Time Sampling (TTS) and Calendar Time Sampling (CTS) schemes. Under the TTS scheme, transactions are sampled by ticks and the full grid consists of all time stamps of observed transactions. Under the CTS scheme time stamps are selected based on calendar time and the previous-tick method is implemented to account for time stamps without observations. For more information of different sampling schemes, including the Business Time Sampling scheme, please refer to Dong and Tse (2014).

actions are often sampled at low sampling frequencies, such as 5 min or 20 min. A drawback of using one subgrid $\mathscr{H}$ at low sampling frequencies is that we lose information given the full grid $\mathscr{G}$. Zhang *et al.* (2005) propose the *subsampling* method to make use of all available observations in $\mathscr{G}$. Given K subgrids, realized volatility RV_K^k are computed in subgrids $\mathscr{G}_K^k$ for $k = 1, \cdots, K$, where

$$\mathscr{G}_K^k = \{t_{k-1},\ t_{k-1+K},\ t_{k-1+2K},\ \cdots,\ t_{k-1+(m_K^k-1)K}\} \tag{4.2.2}$$

and $m_K^k = \max\{z \in \mathbb{N} : t_{k-1+(z-1)K} \le t_n\}$.[4] The integrated volatility is then estimated using all RV_K^k. Thus, the subsampling method makes use of all available observations in $\mathscr{G}$, as $\mathscr{G} = \bigcup_{k=1}^{K} \mathscr{G}_K^k$. For a generic process Z under subgrid $\mathscr{G}_K^k$, we define the intraday return over the interval $[t_{i-}, t_i]$ as $r_{t_i,K,Z} = Z_{t_i} - Z_{t_{i-K}}$, where $t_i \in \mathscr{G}_K^k$.

We now postulate the following assumption for the efficient price process $\{X_t\}$.

Assumption 1. The efficient log-price $\{X_t\}$ follows the Itô process $dX_t = \mu_t dt + \sigma_t dW_t$.

Under assumption 1, the integrated volatility of the price process over $[t_0, t_n]$ is $IV = \int_{t_0}^{t_n} \sigma_s^2 ds$. According to standard arguments, the mean drift of the price process has negligible impact on the analysis of very high-frequency data. Thus, without loss of generality we assume $\mu_t = 0$. For simplicity we do not consider price jumps, although our results are still valid for a general semimartingale process. For example, Jacod and Shiryaev (2003) provide the uniform convergence property of realized volatility to the quadratic variation for the general càdlàg adapted semimartingale process.

[4]Note that $t_{i-} = t_{i-K}$ and $t_{i+} = t_{i+K}$ for any time point $t_i \in \mathscr{G}_K^k$. Higher sampling frequency goes with smaller value of K. When sampling frequency decreases, K increases and m_K^k decreases.

The realized variance (Andersen *et al.* (2001a, 2001b, 2003)) in subgrid $\mathscr{G}_K^k$ is

$$\begin{aligned}
RV_K^k &= \sum_{t_i \in \mathscr{G}_K^k} (Y_{t_i} - Y_{t_{i-}})^2 \\
&= \sum_{t_i \in \mathscr{G}_K^k} \left(r_{t_i,K,X} + \varepsilon_{t_i} - \varepsilon_{t_{i-}}\right)^2 \\
&= \sum_{t_i \in \mathscr{G}_K^k} (r_{t_i,K,X})^2 + \sum_{t_i \in \mathscr{G}_K^k} \left(\varepsilon_{t_i} - \varepsilon_{t_{i-}}\right)^2 + 2 \sum_{t_i \in \mathscr{G}_K^k} r_{t_i,K,X} \left(\varepsilon_{t_i} - \varepsilon_{t_{i-}}\right).
\end{aligned}$$

and the bias of RV_K^k is

$$\mathrm{E}\left[RV_K^k - IV\right] = \mathrm{E}\left[\sum_{t_i \in \mathscr{G}_K^k} \left(\varepsilon_{t_i} - \varepsilon_{t_{i-}}\right)^2\right] + 2\mathrm{E}\left[\sum_{t_i \in \mathscr{G}_K^k} r_{t_i,K,X} \left(\varepsilon_{t_i} - \varepsilon_{t_{i-}}\right)\right]. \quad (4.2.3)$$

The subsampling realized volatility estimate of Zhang *et al.* (2005) is

$$\overline{RV}_K = \frac{1}{K} \sum_{k=1}^{K} RV_K^k, \quad (4.2.4)$$

and the bias of $\overline{RV}_K$ is

$$\mathrm{E}\left[\overline{RV}_K - IV\right] = \frac{1}{K}\mathrm{E}\left[\sum_{t_i \in \mathscr{G}_K^k, k=1,\cdots,K} \left(\varepsilon_{t_i} - \varepsilon_{t_{i-}}\right)^2\right] + \frac{2}{K}\mathrm{E}\left[\sum_{t_i \in \mathscr{G}_K^k, k=1,\cdots,K} r_{t_i,K,X} \left(\varepsilon_{t_i} - \varepsilon_{t_{i-}}\right)\right]. \quad (4.2.5)$$

We further make the following two assumptions about the noise process $\{\varepsilon_t\}$ and the latent log-price $\{X_t\}$.

Assumption 2. The noise process $\{\varepsilon_t\}$ is covariance stationary with mean 0, and auto-covariance function $\pi(s) \equiv E\left(\varepsilon_{t_i}\varepsilon_{t_{i-s}}\right)$ for any $t_i, t_{i-s} \in \mathscr{G}$ and $s \in \mathbb{Z}$.

Assumption 3. The cross-covariance function between the noise process $\{\varepsilon_t\}$ and the latent log-price process $\{X_t\}$ is stationary with $\rho(s,w) \equiv \mathrm{E}\left[\varepsilon_{t_{i+s}}(X_{t_i} - X_{t_{i-w}})\right]$, for $t_{i+s}, t_i, t_{i-w} \in \mathscr{G}$, $s \in 0 \cup \mathbb{N}$ and $w \in \mathbb{N}$. Moreover, the microstructure noise does not predict future latent log-price returns. That is, $\mathrm{E}\left[\varepsilon_{t_{i-w-s}}(X_{t_i} - X_{t_{i-w}})\right] = 0$ for $t_{i-w-s}, t_i, t_{i-w} \in \mathscr{G}$, and $s, w \in \mathbb{N}$.

Based on Assumptions 1-3, from equations (4.2.3) and (4.2.5), we have

$$\mathrm{E}\left[RV_K^k - IV\right] = 2m_K^k\left[\pi(0) - \pi(K) + \rho(0,K)\right] \tag{4.2.6}$$

and

$$\mathrm{E}\left[\overline{RV}_K - IV\right] = 2\overline{m}_K\left[\pi(0) - \pi(K) + \rho(0,K)\right], \tag{4.2.7}$$

where $\overline{m}_K = \sum_{k=1}^{K} m_K^k / K$. By Cauchy-Schwarz inequality, $\pi(0) - \pi(K)$ is always non-negative. Thus, the bias of the realized variance, with or without subsampling, is negative only when $\rho(0,K) < -[\pi(0) - \pi(K)]$. Hence, RV_K^k and $\overline{RV}_K$ underestimate IV only when the cross-correlation between the microstructure noise and latent returns is negative.

Zhou (1996) proposes a subsampling first-order auto-correlation RV estimator to alleviate the bias introduced by the microstructure noise. We denote this estimate by $\overline{RV}_{KA}$, which is defined as

$$\overline{RV}_{KA} = \frac{1}{K}\sum_{k=1}^{K} RV_{KA}^k, \tag{4.2.8}$$

where

$$RV_{KA}^k = RV_K^k + 2A_K^k \tag{4.2.9}$$

and

$$A_K^k = \frac{m_K^k}{m_K^k - 1}\sum_{t_i \in \mathscr{G}_K^k}(Y_{t_{i+}} - Y_{t_i})(Y_{t_i} - Y_{t_{i-}}). \tag{4.2.10}$$

We also define $\overline{A}_K$ as

$$\overline{A}_K = \frac{1}{K}\sum_{k=1}^{K} A_K^k. \tag{4.2.11}$$

Under Assumptions 1-3, the bias of RV_{KA}^{k} is

$$\mathrm{E}\left[RV_{KA}^{k} - IV\right] = 2m_K^k[\pi(K) - \pi(2K) + \rho(K,K)] \tag{4.2.12}$$

and the bias of $\overline{RV}_{KA}$ is

$$\mathrm{E}\left[\overline{RV}_{KA} - IV\right] = 2\overline{m}_K[\pi(K) - \pi(2K) + \rho(K,K)]. \tag{4.2.13}$$

If the noises are iid so that $\pi(s) = 0$ for all $s \neq 0$, the bias of RV_{KA}^{k} reduces to $2m_K^k\rho(K,K)$ and the bias of $\overline{RV}_{KA}$ reduces to $2\overline{m}_K\rho(K,K)$. Furthermore, if the noise process is independent of the latent return, (*i.e.*, $\pi(s) = 0$ for all $s \neq 0$ and $\rho(s,w) = 0$ for all s and w), we have $\mathrm{E}[RV_K^k - IV] = 2m_K^k\pi(0)$ and $\mathrm{E}[RV_{KA}^k - IV] = 0$. As $\mathrm{E}\left[RV_{KA}^k\right] = \mathrm{E}\left[RV_K^k\right] + 2\mathrm{E}\left[A_K^k\right]$, we also have $\mathrm{E}[A_K^k] = -m_K^k\pi(0)$.

4.2.2 Estimating the Noise Variance

Hansen and Lunde (2006) propose to estimate the microstructure noise variance $\pi(0)$ (denoted by ω^2 for simplicity) by

$$\hat{\omega}^2 = \frac{RV_K^k - IV}{2m_K^k}, \tag{4.2.14}$$

They select RV_{KA}^k as an unbiased estimator of IV and use tick-by-tick transactions to compute RV_K^k and RV_{KA}^k in their empirical analysis. In addition, instead of estimating the noise variance at one time scale using RV_K^k, they also calculate the noise variance at two time scales. The noise variance is computed using the difference of the realized variance at two different frequencies K_1 and K_2, with K_2 fixed at 30 min. Thus, they estimate the noise variance by

$$\check{\omega}^2 = \frac{RV_{K_1}^{k_1} - RV_{K_2}^{k_2}}{2(m_{K_1}^{k_1} - m_{K_2}^{k_2})}, \tag{4.2.15}$$

where K_2 is selected to render 30-min sampling frequency ($m_{K_2}^{k_2} = 13$ for the NYSE). Empirically, same as $\hat{\omega}^2$ in equation (4.2.14), they compute $RV_{K_1}^{k_1}$ using tick-by-tick trading prices. On the other hand, Oomen (2006b) and Ubukata and Oya (2009) propose to use the first-order auto-covariance of asset returns to estimate the noise variance as follows

$$\tilde{\omega}^2 = -\frac{A_K^k}{m_K^k}. \tag{4.2.16}$$

It is interesting to note that if RV_{KA}^k is selected as IV in equation (4.2.14), $\hat{\omega}^2$ is identical to $\tilde{\omega}^2$.

One drawback of the noise-variance estimates aforementioned is that there is estimation error when we sample transactions at too high frequencies. We show in our empirical study that the microstructure noise is negatively correlated with high-frequency latent asset returns. The cross correlation between the noise and latent returns will distort the estimates, especially when the magnitude of the noise variance is small. In view of this, we will not use tick-by-tick transactions to calculate $\hat{\omega}^2$, $\breve{\omega}^2$ and $\tilde{\omega}^2$, respectively, in equations (4.2.14)-(4.2.16). However, when we select transactions sparsely and only use transactions in one subgrid, there is heavy loss of information. To deal with this problem, we propose to utilize all available transactions using subsampling method. We modify the noise-variance estimate $\hat{\omega}^2$ in equation (4.2.14) and consider the noise-variance estimate

$$\hat{\omega}^2(K) = \frac{\overline{RV}_K - IV}{2\overline{m}_K}. \tag{4.2.17}$$

For simplicity, we denote this estimate by M1. The following proposition states the property of M1.

Proposition 1. Suppose the efficient log-price $\{X_t\}$ is an Itô process satisfying Assumption 1 and the microstructure noise process is $\{\varepsilon_t\}$, where $\varepsilon_t = Y_t - X_t$. The observed log-price $\{Y_t\}$ is related to $\{X_t\}$ and $\{\varepsilon_t\}$ through $Y_t = X_t + \varepsilon_t$, the noise $\{\varepsilon_t\}$ is iid with $\mathrm{E}(\varepsilon_t) = 0$. Furthermore, assume $\{\varepsilon_t\}$ and $\{X_t\}$ to be independent, with $\mathrm{E}(\varepsilon_t^4) < \infty$. As $n \to \infty$, for M1, we have

$$\sqrt{n}\left(\hat{\omega}^2(K) - \mathrm{E}(\varepsilon_t^2)\middle|X\right) \xrightarrow{L} \mathbf{N}\left(0, \mathrm{E}(\varepsilon_t^4)\right). \tag{4.2.18}$$

The proof of proposition 1 is in the Appendix.[5] Note that empirically we need to use some unbiased and consistent estimate $\widehat{IV}$ to approximate IV for M1.

Furthermore, we also find that there is obvious estimation error when we sample transactions at too low frequencies. We find $\overline{RV}_K$ of a typical NYSE stock decreases when sampling frequency decreases. The mean $\overline{RV}_K$ at low sampling frequency is significantly smaller than the mean Realized Kernel (RK) estimate, where the RK method is noise-robust. Thus, we do not recommend using $\overline{RV}_K$ at low sampling frequency, such as 30-min, to investigate the noise property. Instead, we suggest to estimate the noise variance using difference of $\overline{RV}_{K_i}$, for $i = 1,2$, with $K_1 \neq K_2$, at moderately high sampling frequencies, such as 1 min or 2 min. This will helps to alleviate problems introduced by either too high or too low sampling frequencies. We modify the noise-variance estimate $\breve{\omega}^2$ in equations (4.2.15) and consider the following noise-variance estimate

$$\breve{\omega}^2(K_1,K_2) = \frac{\overline{RV}_{K_1} - \overline{RV}_{K_2}}{2(\overline{m}_{K_1} - \overline{m}_{K_2})}, \tag{4.2.19}$$

where we do not necessarily select K_i, for $i = 1,2$, at one tick sampling frequency or low sampling frequency such as 30 min. For simplicity, we denote this estimate by M2.[6] The following proposition states the properties of M2.

Proposition 2. Under the same assumptions as in Proposition 1. As $n \to \infty$, for M2 with $K_1 \neq K_2$, we have

$$\sqrt{n}\left(\breve{\omega}^2(K_1,K_2) - \mathrm{E}(\varepsilon_t^2)\middle|X\right) \xrightarrow{L} \mathbf{N}\left(0, \mathrm{E}(\varepsilon_t^4) + \frac{2K_1K_2}{(K_2-K_1)^2}\left(\mathrm{E}(\varepsilon_t^2)\right)^2\right). \tag{4.2.20}$$

[5] We treat K as fixed in the paper.

[6] Zhang *et al.* (2005) also suggest to estimate the noise variance using M2, but they fix K_1 at one tick sampling frequency.

The proof of proposition 2 is in the Appendix. Note that theoretically M2 always report larger estimation variance than M1. However, in implementation, we need to select an unbiased IV estimate $\widehat{IV}$ for M1, and the estimation error introduced by $\widehat{IV}$ will further distort the estimated noise variance. In contrast, for M2, the latent integrated volatility is included implicitly in $\overline{RV}_{K_2}$ and the estimation error of $\overline{RV}_{K_i}$, for $i = 1, 2$, may even be mitigated when we do the subtraction. We investigate M1 and M2 in our Monte Carlo and empirical study and show the better performance of M2 than M1.

We further modify $\tilde{\omega}^2$ in equation (4.2.16) using the subsampling method and estimate the noise variance via

$$\tilde{\omega}^2(K) = -\frac{\overline{A}_K}{\overline{m}_K}. \tag{4.2.21}$$

We denote this estimate by M3. The properties of M3 can be found in Section 3.2 of Ubukata and Oya (2009).

Instead of estimating the noise variance only at one or two time scales using M1 and M2, we may estimate the noise variance using subsampling realized variance at multiple time scales. Now we compute the noise variance based on subsampling realized variance at sampling frequencies centering around K or K_i, for $i = 1, 2$. That is, we estimate $\overline{RV}_{K+lq}$, for $l = -L, \cdots, L$, with $L, q \in \mathbb{N}$,[7] and define the weighted average of subsampling realized variance as

$$\overline{RV}_{K,L,q} = \sum_{l=-L}^{L} \frac{(K+lq)\overline{RV}_{K+lq}}{(2L+1)K}. \tag{4.2.22}$$

The new noise-variance estimator is then given by

$$\hat{\omega}^2(K,L,q) = \frac{\overline{RV}_{K,L,q} - \widehat{IV}}{2\overline{m}_K}. \tag{4.2.23}$$

We call this estimate M1A. The following proposition states the properties of M1A.

[7] In the paper, we use $q = 1$ tick under the TTS scheme and $q = 10$ sec under the CTS scheme.

Proposition 3. Under the same assumptions as in Proposition 1, as $n \to \infty$, for M1A, we have

$$\sqrt{n}\left(\hat{\omega}^2(K,L,q) - \mathrm{E}(\varepsilon_t^2)\middle|X\right) \xrightarrow{L} \mathbf{N}\left(0, \mathrm{E}(\varepsilon_t^4) - \frac{2L}{2L+1}\left(\mathrm{E}(\varepsilon_t^2)\right)^2\right). \tag{4.2.24}$$

Compared with M1, M1A has smaller variance. Furthermore, we can calculate the noise variance using

$$\breve{\omega}^2(K_1,K_2,L,q) = \frac{\overline{RV}_{K_1,L,q} - \overline{RV}_{K_2,L,q}}{2(\overline{m}_{K_1} - \overline{m}_{K_2})}. \tag{4.2.25}$$

For simplicity, we call this estimate M2A. The following proposition states the properties of M2A.

Proposition 4. Under the same assumptions as in Proposition 1, as $n \to \infty$, for M2A with $K_1 \neq K_2$, we have

$$\begin{aligned} &\sqrt{n}\left(\breve{\omega}^2(K_1,K_2,L,q) - \mathrm{E}(\varepsilon_t^2)\middle|X\right) \\ &\xrightarrow{L} \mathbf{N}\left(0, \mathrm{E}(\varepsilon_t^4) + \frac{1}{2L+1}\left(-2L + \frac{2K_1K_2}{(K_2-K_1)^2}\right)\left(\mathrm{E}(\varepsilon_t^2)\right)^2\right). \end{aligned} \tag{4.2.26}$$

The proof of the propositions 3 and 4 can be found in the Appendix.

Asymptotically, compared with M2, M2A has smaller error variance for all $L > 0$. Although theoretically M1A has smaller error variance than M2A, as we discussed before, empirically $\widehat{IV}$ may distort M1A. We investigate the performances of M1A and M2A in our Monte Carlo and empirical study and show the superiority of M2A. Our noise-variance estimates M1A and M2A share some similarities to the multi-scale approach of Zhang (2006), although we use different weighting functions here. Zhang's (2006) estimator has the form $\widehat{\langle X,X\rangle}^{(n)} = \sum_{i=1}^{M} \alpha_i \overline{RV}_{K_i}$ where $\{\alpha_i\}, i = 1,\cdots,M$, are selected so that $\widehat{\langle X,X\rangle}^{(n)}$ is asymptotically unbiased and has optimal convergence rate of $n^{-1/4}$, with

$$\sum_{i=1}^{M} \alpha_i = 1 \quad \text{and} \quad \sum_{i=1}^{M} \frac{\alpha_i(n+1-K_i)}{K_i} = 0. \tag{4.2.27}$$

Here $\{\alpha_i\}, i = 1, \cdots, M$, take negative values for some i. In contrast, the weights in our estimates are always positive. While the MSRV method is proposed to estimate the realized volatility with optimal convergence rate of $n^{-1/4}$, our estimators are proposed to estimate the noise variance with smaller estimation error.

We list all five noise-variance estimates in Table 1 for ease of comparison.

4.3 Characterizing Noise Properties using Volatility Signature Plot

Bandi and Russell (2006) point out that the noise associated with transactions is quite different from noise associated with quotes, due to their different formation mechanisms. Different markets, sampling methodologies and price measurements may render different noise properties as well. In this paper we investigate the properties of the noise associated with the NYSE transaction prices. We use the subsampling realized variance $\overline{RV}_K$ and subsampling first-order auto-correlation bias-corrected realized variance $\overline{RV}_{KA}$ to calculate the volatility signature plots. We use the NYSE transaction data from January 2010 to April 2013 of the top 40 market-capitalization stocks (as of 2010) with data compiled from the WRDS TAQ database. To clean the raw data, we follow the steps described in Tse and Dong (2014). When there are multiple transactions at the same time stamp, we use the average price weighted by trading volume.

In Figure 1, we present the volatility signature plot using the mean $\overline{RV}_K$ averaged over all trading days for stock JPM.[8] Under the TTS scheme, the number of subsamples K is determined such that at each trading day the mean sampling duration of each subgrid is approximately equal to the values specified in the x-axis of the figure.[9] We compute the daily volatility using the noise-robust Realized Kernel

[8]To save space we only provide calculated results of JPM under the TTS scheme. Results for other stocks as well as under the CTS scheme are available in our online Appendix.

[9]Based on this setting, the number of subsamples K will vary across trading days for a fixed value in the x-axis. For example, the average transaction duration may be 5 sec in one day and 20

(RK) method of Barndorff-Nielsen *et al.* (2008) and plot mean RK estimate in the figure.[10] The confidence interval of the volatility level is also provided using the method described in Hansen and Lunde (2006) but we use RK as the conditionally unbiased estimate of IV.

Figure 1 shows the impact of different sampling frequencies on $\overline{RV}_K$. When the sampling frequency is high, (*i.e.*, small average transaction duration and small number of subgrids), mean $\overline{RV}_K$ significantly deviates from mean RK.[11] This coincides with the findings in the literature that the microstructure noise distorts the realized variance estimate at high sampling frequency. However, we observe even more significant deviation of mean $\overline{RV}_K$ from mean RK at low sampling frequency. When the average transaction duration is above 20 min, $\overline{RV}_K$ lies below the lower volatility confidence bound for *all* stocks investigated. This is in contrast with results in the literature that the realized volatility stabilizes at low sampling frequency because of the mitigated influence of noise variance. Thus, selecting the sampling frequency is an important issue when $\overline{RV}_K$ is used to estimate the microstructure-noise variance. A very small variation of $\overline{RV}_K$ will result in very different NSR estimates.[12]

We further report volatility signature plots of $\overline{RV}_K$ and $\overline{RV}_{KA}$ averaged over all trading days for JPM in Figure 2, focusing on sampling frequencies below 5 min.[13] At ultra high sampling frequencies, for 30 out of 40 stocks $\overline{RV}_K$ increases

sec in another. At the 1-min target average sampling duration, we need to sample every 12 ticks and every 3 ticks separately on these two days. Sensitivity of $\overline{RV}_K$ to the sampling frequency in our latter analysis tends to support this setting under the TTS scheme, which is superior to sampling transactions by the same number of ticks over all trading days.

[10] Barndorff-Nielsen *et al.* (2009) suggest to use $\overline{RV}_K$ at 15-min sampling frequency to calculate the bandwidth of the realized kernel. In contrast, based on the results in this section, throughout this paper we use $\overline{RV}_K$ at 3-min sampling frequency under the TTS scheme to calculate the RK bandwidth.

[11] Specifically, $\overline{RV}_K$ of 8 stocks cross the upper confidence bound and $\overline{RV}_K$ of 3 stocks cross the lower confidence bound at 5 sec sampling frequency for the 40 stocks under the TTS scheme.

[12] Empirically when we sample transactions at low sampling frequencies, the sampled transactions at subgrid $\mathscr{H} = \{t_1, \cdots, t_h\}$ may not cover 6.5 hours. We adjust this problem by multiplying RV_K^k to $23400/(t_h - t_1)$ at each subgrid $\mathscr{G}_K^k$ and then compute the adjusted subsampling realized variance $\overline{RV}_{KA^*}$. The $\overline{RV}_{KA^*}$ volatility signature plots in Figure 1 suggest similar results, although the difference between $\overline{RV}_{KA^*}$ and RK is smaller.

[13] We only provide $\overline{RV}_K$ and $\overline{RV}_{KA}$ of JPM under the TTS scheme. Results for other stocks as well

as the sampling frequency decreases and for more than half of the stocks mean $\overline{RV}_K$ is smaller than mean RK. This provides direct empirical evidence of the negative cross-correlation between the microstructure noise and the latent transaction returns. The negative correlation is also suggested by Hansen and Lunde (2006) but they use quotation data and cointegration analysis. Although $\overline{RV}_{KA}$ varies a lot across different stocks at ultra high sampling frequencies, they stabilize after 1-min sampling frequency. Specifically, at 1-min sampling frequency mean $\overline{RV}_{KA}$ are often very close to the mean RK. Thus, we treat the microstructure noise as iid when the sampling frequency is lower than 1-min.

4.4 Simulation Study

We conduct a Monte Carlo (MC) study to compare the performances of all five microstructure noise-variance estimates M1, M1A, M2, M2A and M3 across different sampling frequencies.

4.4.1 Simulation Setup

The setup of our simulation is similar to that implemented in Tse and Yang (2012), with minor modifications. We assume the following price generation process (Heston (1993)):

$$d\ \log X_t = \left(\mu - \frac{\sigma_t^2}{2}\right) dt + \sigma_t dW_{1t},$$

$$d\sigma_t^2 = \kappa(\alpha - \sigma_t^2)dt + \gamma\sigma_t dW_{2t},$$

with $\mu = 0.05$, $\kappa = 5$, $\alpha = 0.04$ (the long run annualized daily standard deviation is around 20%), and $\gamma = 0.5$. The correlation coefficient between the two Brownian motions W_{1t} and W_{2t} equals to -0.5. We generate simulated transactions with initial value of $X_0 = \log(60)$ and $\sigma_0 = 10\%$. We introduce sparsity to the simulated trans-

as under the CTS scheme are available in our online Appendix.

actions and consider three cases for which the average transaction duration (sparsity parameter) equals 5 sec, 10 sec and 20 sec. To obtain the sparse transactions, we first simulate transactions sec by sec and irregularly spaced transactions are then obtained based on an exponential distribution with mean equal to the sparsity parameter. We consider iid microstructure noise with constant NSR, with NSR = 0, 0.001%, 0.01% and 0.1%. A price rounding error of 0.01 is introduced to the simulated price process. We do 1000 simulation runs for each model with 1000 trading days for each simulation run.

4.4.2 Simulation Results

For M2 and M2A, we select K_1 at 1-min sampling frequency and select K_2 at 2-min, 3-min, 5-min, 10-min and 30-min sampling frequencies. For M1, M1A and M3, we estimate the noise variance at 1-min, 3-min, 5-min, 10-min and 30-min sampling frequencies. For M1 and M1A, we use RK as an unbiased estimate of IV. For M1A and M2A, we use $L = 2$, $q = 1$ tick under the TTS scheme and $L = 2$, $q = 10$ sec under the CTS scheme. NSR is calculated at each simulation run based on the estimated noise variance and the true daily integrated volatility. We use $\overline{RV}_{KA^*}$ instead of $\overline{RV}_K$ to calculate the noise variance. Tables 2-6 report the ME and RMSE of the NSR estimates using simulated prices with different sparsity and NSR parameters under the CTS and TTS schemes. NSR in the first columns of the tables are the target NSR fixed for the simulated noise process. True NSR in the second column are the realized NSR incorporating the price rounding error. One cent price rounding error boosts NSR up by 0.0016%-0.0019%.

Among all five noise variance estimates, M2 and M2A perform the best in reporting ME and RMSE closest to zero across all sampling frequencies investigated. They are also more robust than M1, M1A and M3 across different sampling frequencies, and their superiority is especially significant at low sampling frequencies. High sampling frequency is preferred, as long as the iid noise assumption is satisfied. M2A performs better than M2 with generally smaller RMSE, which is consistent with the propositions in Section 2. Under the CTS scheme, M2 sometimes report smaller RMSE in comparison to M2A when transactions are more sparse and NSR is big. This may be due to errors introduced by the previous-tick method. We also estimate the noise variance using $\overline{RV}_K$ instead of $\overline{RV}_{KA^*}$ for M1, M1A, M2 and M2A and find that all estimates exhibit obvious bias, as well as larger RMSE.[14] In sum, our results suggest that M2 and M2A with $\overline{RV}_{KA^*}$ provide the best estimates of the market microstructure noise variance.

We present the $\overline{RV}_K$ and $\overline{RV}_{KA}$ volatility signature plots of one simulation run under the CTS and TTS schemes in Figure 3 with a sparsity parameter of 10 sec and NSR of 0.01%.[15] The mean true integrated volatility and mean RK are provided in the figure as well. We construct the confidence band based on true daily integrated volatility. Under the TTS scheme, $\overline{RV}_{KA}$ performs very well in getting rid of the noise effect in volatility estimation. However, $\overline{RV}_{KA}$ under the CTS scheme sometimes lie above the upper confidence bound. This suggests that the previous-tick method may introduce noise that is autocorrelated or cross-correlated with the latent returns. The autocorrelation of noise is more obvious when the NSR parameter or the sparsity parameter is big. In contrast, the noise introduced by the price rounding error tends to be independent. Mean $\overline{RV}_K$ is very close to mean RK at all sampling frequencies in our simulation study. This is in contrast to the findings in our empirical analysis, which suggests that the Heston model may not describe the empirical trading prices very well.

4.5 Empirical Estimates of Noise-to-Signal Ratio

We estimate NSR of the 40 NYSE stocks using M2 and M2A and compare their results against the M1, M1A and M3 methods. We calculate NSR using M2 and M2A at 1-min versus 2-min sampling frequencies, and M1, M1A and M3 at

[14]Results are available in the online Appendix.

[15]We present volatility signature plots of all cases in the online Appendix.

1-min sampling frequency. For M1, M1A, M2 and M2A, we use $\overline{RV}_{KA^*}$ instead of $\overline{RV}_K$ to calculate the noise variance, with all estimates implemented under the TTS scheme. For M1 and M1A, we use RK as the unbiased estimate of IV. The results are reported in Table 7. Using M2 and M2A, we find all stocks have very small NSR, except for MMM, which reports negative figures. Specifically, 32 out of 39 stocks report NSR with positive values below 0.006% and only C and BAC report NSR with values above 0.01%. The estimated NSR is much smaller than that reported in the literature. For example, Hansen and Lunde (2006) report that NSR is slightly below 0.1%. M2A reports smaller NSR than M2 for 37 out of 39 stocks (MMM is not take into account). NSR estimated using M2 and M2A are quite different from those obtained using M1, M1A and M3.

We further check the robustness of all five noise-variance estimates by implementing them at different sampling frequencies. We calculate NSR using M2 and M2A at 1-min versus 3-min sampling frequencies, and M1, M1A and M3 at 2-min sampling frequency. The results are reported in Table 8. For M2 and M2A, NSR increases slightly and there are still 29 (30) out of 40 stocks for M2 (M2A) with NSR below 0.006%. Only C, BAC and PG report NSR with values above 0.01%. In contrast, NSR calculated from M1, M1A and M3 change significantly across different sampling frequencies. Only 3 (4) stocks report negative NSR for M1 (M1A) at 1-min sampling frequency but 20 (21) out of 40 stocks report negative NSR at 2-min sampling frequency. For M3, 29 out of 40 stocks report negative NSR at 1-min sampling frequency but only 3 stocks report negative NSR at 2-min sampling frequency. We also estimate NSR using M2 and M2A at 1-min versus 30-min sampling frequency and using M1, M1A and M3 at 30-min sampling frequencies.[16] All NSR estimates are very different from the results in Tables 7 and 8. Specifically, the estimated NSR generally increase by more than 5 times for M2 and M2A, and increase or decrease by more than 50 times for M1, M1A and M3. This finding is interesting since in our simulation study all noise variance estimates report very

[16]We report these results in our online Appendix.

good unbiasedness property across different sampling frequencies. Finally, we calculate NSR using M2A at 1-min versus 2-min sampling frequencies by every 100 trading days. We report mean NSR by averaging across all 40 stocks in Figure 4. NSR tends to vary over time and is quite high during the period around 2010.

4.6 Conclusion

We investigate the noise properties of NYSE high-frequency trading prices. Selecting the sampling frequency of the subsampled realized variance $\overline{RV}_K$ is important for the noise-variance estimation. For the selected NYSE stocks in 2010 to 2013, the microstructure noise is negatively correlated with the latent asset returns at high sampling frequencies, while at low sampling frequencies $\overline{RV}_K$ always decrease when the sampling frequency decreases. Specifically, $\overline{RV}_K$ at low sampling frequencies, such as 20 min, is significantly different from the Realized Kernel estimate. This observation is largely neglected in the literature, leading researchers to favor using low sampling frequency to alleviate the influence of noise.

We propose two noise-variance estimates M2 and M2A using the difference of subsampled realized variance at two or multiple time scales. We compare five noise-variance estimates using Monte Carlo study and our proposed estimates perform the best in reporting lower ME and RMSE. To estimate noise variance, high sampling frequency is preferred whenever the noise can be recognized as iid. In practice, we suggest sampling transactions at moderately high frequencies, such as 1 min or 2 min. We estimate NSR using transaction data for NYSE stocks. The estimated NSR of the selected stocks is around 0.005% from 2010 to 2013. The magnitude of NSR is so small that it is delicate to investigate the microstructure noise properties. For example, we cannot ignore the cross correlation between the microstructure noise and latent asset returns when investigating the auto-covariance of the microstructure noise.

Table 4.1: Summary of Noise-variance Estimators

Model	Notation	Noise-variance Estimator	Description of Estimator
M1	$\hat{\omega}^2(K)$	$\dfrac{\overline{RV}_K - \widehat{IV}}{2\bar{m}_K}$	$\overline{RV}_K = \frac{1}{K}\sum_{k=1}^{K} RV_K^k$, $RV_K^k = \sum_{t_i \in \mathcal{G}_K^k} (Y_{t_i} - Y_{t_{i-}})^2$, $\bar{m}_K = \sum_{k=1}^{K} m_K^k / K$ and $m_K^k = \lvert\mathcal{G}_K^k\rvert$. We use the Realized Kernel method as the unbiased estimate of IV.
M1A	$\hat{\omega}^2(K_1, L, q)$	$\dfrac{\overline{RV}_{K,L,q} - \widehat{IV}}{2\bar{m}_K}$	$\overline{RV}_{K,L,q} = \sum_{l=-L}^{L} \frac{(K+lq)\overline{RV}_{K+lq}}{(2L+1)K}$, and $\bar{RV}_K$, $\bar{m}_K$ same as M1.
M2	$\check{\omega}^2(K_1, K_2)$	$\dfrac{\overline{RV}_{K_1} - \overline{RV}_{K_2}}{2(\bar{m}_{K_1} - \bar{m}_{K_2})}$	See M1.
M2A	$\check{\omega}^2(K_1, K_2, L, q)$	$\dfrac{\overline{RV}_{K_1,L,q} - \overline{RV}_{K_2,L,q}}{2(\bar{m}_{K_1} - \bar{m}_{K_2})}$	See M1A.
M3	$\tilde{\omega}^2(K)$	$-\dfrac{\overline{A}_K}{\bar{m}_K}$	$\overline{A}_K = \frac{1}{K}\sum_{k=1}^{K} A_K^k$, $A_K^k = \frac{m_K^k}{m_K^k - 1}\sum_{t_i \in \mathcal{G}_K^k} (Y_{t_{i+}} - Y_{t_i})(Y_{t_i} - Y_{t_{i-}})$ and $\bar{m}_K$ same as M1

Table 4.2: ME and RMSE for NSR estimates M1 using $\overline{RV}_{KA^*}$ under the CTS and TTS schemes

Scheme	Sparsity	NSR	True NSR	ME ($\times 10^4$)					RMSE ($\times 10^4$)				
			($\times 10^4$)	1-min	3-min	5-min	10-min	30-min	1-min	3-min	5-min	10-min	30-min
CTS	5-sec	0%	0.1580	0.0220	0.0701	0.1167	0.2289	0.5304	0.0464	0.0801	0.1706	0.6692	4.4869
		0.001%	0.2583	0.0194	0.0622	0.1037	0.2024	0.4532	0.0453	0.0731	0.1617	0.6605	4.4771
		0.01%	1.1594	-0.0043	-0.0093	-0.0153	-0.0358	-0.2613	0.0418	0.0399	0.1237	0.6270	4.4490
		0.1%	9.9994	-0.2257	-0.6745	-1.1223	-2.2515	-6.9025	0.2303	0.6760	1.1287	2.3327	8.1586
	10-sec	0%	0.1581	0.0285	0.0892	0.1465	0.2901	0.7177	0.0594	0.1154	0.1714	0.6192	4.4069
		0.001%	0.2584	0.0253	0.0802	0.1314	0.2601	0.6267	0.0578	0.1088	0.1587	0.6056	4.3920
		0.01%	1.1597	-0.0037	-0.0007	-0.0035	-0.0093	-0.1788	0.0525	0.0743	0.0900	0.5454	4.3389
		0.1%	9.9805	-0.2738	-0.7468	-1.2475	-2.4974	-7.6380	0.2798	0.7514	1.2512	2.5536	8.7369
	20-sec	0%	0.1582	0.0359	0.1348	0.2238	0.4459	1.1817	0.0735	0.1820	0.2549	0.6260	4.2743
		0.001%	0.2587	0.0275	0.1247	0.2072	0.4120	1.0800	0.0699	0.1748	0.2406	0.6018	4.2450
		0.01%	1.1609	-0.0470	0.0358	0.0584	0.1137	0.1831	0.0797	0.1284	0.1364	0.4507	4.0940
		0.1%	9.9717	-0.7589	-0.7831	-1.3040	-2.6125	-7.9985	0.7622	0.7943	1.3113	2.6478	8.9421
TTS	5-sec	0%	0.1580	0.0222	0.0669	0.1138	0.2258	0.5394	0.0456	0.0704	0.1637	0.6677	4.4912
		0.001%	0.2583	0.0195	0.0590	0.1008	0.1995	0.4620	0.0444	0.0629	0.1547	0.6590	4.4807
		0.01%	1.1594	-0.0046	-0.0128	-0.0186	-0.0394	-0.2537	0.0403	0.0255	0.1176	0.6266	4.4505
		0.1%	9.9994	-0.2288	-0.6782	-1.1264	-2.2558	-6.8973	0.2327	0.6787	1.1318	2.3367	8.1539
	10-sec	0%	0.1581	0.0301	0.0885	0.1438	0.2811	0.7179	0.0594	0.1045	0.1573	0.6204	4.4175
		0.001%	0.2584	0.0271	0.0795	0.1288	0.2513	0.6270	0.0578	0.0970	0.1436	0.6073	4.4028
		0.01%	1.1597	-0.0011	-0.0013	-0.0057	-0.0172	-0.1792	0.0515	0.0561	0.0629	0.5509	4.3486
		0.1%	9.9805	-0.2613	-0.7469	-1.2492	-2.5041	-7.6428	0.2671	0.7494	1.2507	2.5607	8.7438
	20-sec	0%	0.1582	0.0456	0.1359	0.2204	0.4348	1.1683	0.0787	0.1706	0.2273	0.6011	4.2638
		0.001%	0.2587	0.0421	0.1257	0.2034	0.4007	1.0663	0.0767	0.1627	0.2108	0.5765	4.2350
		0.01%	1.1609	0.0115	0.0353	0.0542	0.1016	0.1695	0.0657	0.1096	0.0775	0.4244	4.0882
		0.1%	9.9717	-0.2680	-0.7912	-1.3107	-2.6272	-8.0130	0.2769	0.7987	1.3121	2.6573	8.9529

Table 4.3: ME and RMSE for NSR estimates M1A using $\overline{RV}_{KA^*}$ under the CTS and TTS schemes

Scheme	Sparsity	NSR	True NSR	ME ($\times 10^4$)					RMSE ($\times 10^4$)				
			($\times 10^4$)	1-min	3-min	5-min	10-min	30-min	1-min	3-min	5-min	10-min	30-min
CTS	5-sec	0%	0.1580	0.0221	0.0702	0.1167	0.2289	0.5307	0.0454	0.0799	0.1705	0.6693	4.4882
		0.001%	0.2583	0.0194	0.0622	0.1037	0.2024	0.4536	0.0443	0.0729	0.1616	0.6606	4.4783
		0.01%	1.1594	-0.0044	-0.0093	-0.0154	-0.0358	-0.2609	0.0405	0.0396	0.1236	0.6271	4.4500
		0.1%	9.9994	-0.2265	-0.6744	-1.1226	-2.2515	-6.9023	0.2307	0.6759	1.1289	2.3326	8.1590
	10-sec	0%	0.1581	0.0279	0.0892	0.1466	0.2901	0.7179	0.0582	0.1149	0.1714	0.6193	4.4080
		0.001%	0.2584	0.0244	0.0802	0.1315	0.2601	0.6270	0.0564	0.1083	0.1588	0.6057	4.3931
		0.01%	1.1597	-0.0076	-0.0007	-0.0034	-0.0093	-0.1785	0.0518	0.0735	0.0899	0.5456	4.3399
		0.1%	9.9805	-0.3065	-0.7470	-1.2472	-2.4973	-7.6377	0.3116	0.7515	1.2509	2.5535	8.7372
	20-sec	0%	0.1582	0.0338	0.1348	0.2238	0.4459	1.1820	0.0719	0.1814	0.2547	0.6261	4.2752
		0.001%	0.2587	0.0241	0.1247	0.2072	0.4120	1.0803	0.0679	0.1742	0.2404	0.6019	4.2459
		0.01%	1.1609	-0.0624	0.0358	0.0585	0.1136	0.1833	0.0891	0.1276	0.1360	0.4508	4.0949
		0.1%	9.9717	-0.8920	-0.7835	-1.3039	-2.6126	-7.9984	0.8946	0.7945	1.3112	2.6479	8.9423
TTS	5-sec	0%	0.1580	0.0222	0.0670	0.1138	0.2259	0.5395	0.0453	0.0704	0.1638	0.6678	4.4917
		0.001%	0.2583	0.0195	0.0591	0.1008	0.1995	0.4620	0.0440	0.0629	0.1547	0.6591	4.4812
		0.01%	1.1594	-0.0046	-0.0126	-0.0186	-0.0394	-0.2536	0.0400	0.0251	0.1176	0.6267	4.4510
		0.1%	9.9994	-0.2287	-0.6777	-1.1263	-2.2557	-6.8969	0.2324	0.6782	1.1316	2.3365	8.1539
	10-sec	0%	0.1581	0.0301	0.0884	0.1438	0.2811	0.7186	0.0583	0.1039	0.1573	0.6203	4.4189
		0.001%	0.2584	0.0270	0.0794	0.1287	0.2513	0.6277	0.0567	0.0964	0.1436	0.6073	4.4042
		0.01%	1.1597	-0.0012	-0.0013	-0.0058	-0.0172	-0.1785	0.0502	0.0552	0.0630	0.5508	4.3499
		0.1%	9.9805	-0.2614	-0.7469	-1.2496	-2.5041	-7.6422	0.2668	0.7493	1.2510	2.5606	8.7440
	20-sec	0%	0.1582	0.0454	0.1357	0.2206	0.4347	1.1681	0.0752	0.1684	0.2270	0.6013	4.2651
		0.001%	0.2587	0.0420	0.1255	0.2036	0.4007	1.0661	0.0732	0.1603	0.2105	0.5767	4.2364
		0.01%	1.1609	0.0116	0.0352	0.0543	0.1016	0.1694	0.0613	0.1061	0.0760	0.4247	4.0898
		0.1%	9.9717	-0.2683	-0.7913	-1.3103	-2.6269	-8.0129	0.2756	0.7982	1.3116	2.6569	8.9535

Table 4.4: ME and RMSE for NSR estimates M2 using $\overline{RV}_{KA^*}$ under the CTS and TTS schemes

Scheme	Sparsity	NSR	True NSR	ME ($\times 10^4$)					RMSE ($\times 10^4$)				
			($\times 10^4$)	2-min	3-min	5-min	10-min	30-min	2-min	3-min	5-min	10-min	30-min
CTS	5-sec	0%	0.1580	-0.0014	-0.0020	-0.0016	-0.0009	0.0045	0.0410	0.0512	0.0672	0.0944	0.1671
		0.001%	0.2583	-0.0014	-0.0020	-0.0017	-0.0009	0.0045	0.0410	0.0513	0.0673	0.0946	0.1673
		0.01%	1.1594	-0.0012	-0.0018	-0.0015	-0.0008	0.0046	0.0417	0.0521	0.0679	0.0949	0.1672
		0.1%	9.9994	-0.0004	-0.0013	-0.0015	-0.0006	0.0046	0.0495	0.0574	0.0715	0.0967	0.1659
	10-sec	0%	0.1581	-0.0015	-0.0019	-0.0010	-0.0006	0.0047	0.0414	0.0517	0.0673	0.0942	0.1674
		0.001%	0.2584	-0.0019	-0.0021	-0.0012	-0.0008	0.0046	0.0414	0.0514	0.0671	0.0940	0.1674
		0.01%	1.1597	-0.0060	-0.0053	-0.0038	-0.0031	0.0023	0.0429	0.0522	0.0674	0.0941	0.1670
		0.1%	9.9805	-0.0485	-0.0373	-0.0304	-0.0267	-0.0199	0.0741	0.0706	0.0782	0.0996	0.1665
	20-sec	0%	0.1582	-0.0168	-0.0136	-0.0111	-0.0097	-0.0036	0.0439	0.0525	0.0670	0.0942	0.1668
		0.001%	0.2587	-0.0265	-0.0211	-0.0174	-0.0152	-0.0088	0.0486	0.0549	0.0684	0.0948	0.1670
		0.01%	1.1609	-0.1140	-0.0883	-0.0733	-0.0648	-0.0549	0.1213	0.1020	0.0988	0.1135	0.1750
		0.1%	9.9717	-0.9699	-0.7468	-0.6227	-0.5530	-0.5093	0.9717	0.7494	0.6269	0.5611	0.5350
TTS	5-sec	0%	0.1580	-0.0000	-0.0004	-0.0009	-0.0006	0.0043	0.0416	0.0520	0.0673	0.0949	0.1682
		0.001%	0.2583	-0.0002	-0.0005	-0.0010	-0.0007	0.0041	0.0416	0.0520	0.0673	0.0950	0.1683
		0.01%	1.1594	-0.0002	-0.0004	-0.0010	-0.0006	0.0042	0.0418	0.0522	0.0674	0.0948	0.1679
		0.1%	9.9994	-0.0010	-0.0005	-0.0013	-0.0008	0.0038	0.0464	0.0543	0.0685	0.0945	0.1654
	10-sec	0%	0.1581	-0.0006	-0.0011	0.0001	0.0009	0.0052	0.0440	0.0541	0.0696	0.0986	0.1748
		0.001%	0.2584	-0.0003	-0.0010	0.0002	0.0009	0.0053	0.0438	0.0539	0.0695	0.0986	0.1747
		0.01%	1.1597	-0.0004	-0.0011	0.0001	0.0008	0.0052	0.0447	0.0545	0.0697	0.0987	0.1744
		0.1%	9.9805	-0.0004	-0.0012	-0.0003	0.0004	0.0050	0.0549	0.0598	0.0725	0.0997	0.1721
	20-sec	0%	0.1582	-0.0009	-0.0005	0.0005	0.0011	0.0057	0.0459	0.0548	0.0698	0.0966	0.1708
		0.001%	0.2587	-0.0011	-0.0006	0.0005	0.0011	0.0057	0.0460	0.0547	0.0699	0.0968	0.1709
		0.01%	1.1609	-0.0010	-0.0006	0.0005	0.0012	0.0058	0.0475	0.0556	0.0704	0.0970	0.1706
		0.1%	9.9717	-0.0002	-0.0004	0.0006	0.0013	0.0058	0.0646	0.0648	0.0762	0.0996	0.1695

Table 4.5: ME and RMSE for NSR estimates M2A using $\overline{RV}_{KA^*}$ under the CTS and TTS schemes

Scheme	Sparsity	NSR	True NSR	ME ($\times 10^4$)					RMSE ($\times 10^4$)				
			($\times 10^4$)	2-min	3-min	5-min	10-min	30-min	2-min	3-min	5-min	10-min	30-min
CTS	5-sec	0%	0.1580	-0.0015	-0.0020	-0.0016	-0.0009	0.0045	0.0382	0.0494	0.0659	0.0936	0.1667
		0.001%	0.2583	-0.0014	-0.0019	-0.0016	-0.0009	0.0045	0.0382	0.0495	0.0660	0.0937	0.1669
		0.01%	1.1594	-0.0015	-0.0019	-0.0016	-0.0009	0.0045	0.0387	0.0500	0.0664	0.0939	0.1667
		0.1%	9.9994	-0.0022	-0.0025	-0.0024	-0.0015	0.0037	0.0436	0.0537	0.0692	0.0952	0.1652
	10-sec	0%	0.1581	-0.0028	-0.0027	-0.0018	-0.0012	0.0041	0.0386	0.0499	0.0660	0.0934	0.1670
		0.001%	0.2584	-0.0038	-0.0035	-0.0024	-0.0018	0.0036	0.0386	0.0496	0.0658	0.0933	0.1670
		0.01%	1.1597	-0.0138	-0.0110	-0.0086	-0.0074	-0.0017	0.0417	0.0512	0.0665	0.0935	0.1666
		0.1%	9.9805	-0.1141	-0.0863	-0.0713	-0.0631	-0.0537	0.1244	0.1029	0.0997	0.1136	0.1733
	20-sec	0%	0.1582	-0.0209	-0.0167	-0.0137	-0.0120	-0.0058	0.0438	0.0520	0.0665	0.0938	0.1666
		0.001%	0.2587	-0.0332	-0.0262	-0.0217	-0.0190	-0.0123	0.0508	0.0557	0.0686	0.0948	0.1669
		0.01%	1.1609	-0.1442	-0.1115	-0.0926	-0.0820	-0.0709	0.1494	0.1219	0.1132	0.1236	0.1804
		0.1%	9.9717	-1.2295	-0.9462	-0.7890	-0.7009	-0.6470	1.2306	0.9480	0.7921	0.7070	0.6672
TTS	5-sec	0%	0.1580	-0.0002	-0.0005	-0.0010	-0.0006	0.0042	0.0408	0.0514	0.0669	0.0947	0.1682
		0.001%	0.2583	-0.0003	-0.0006	-0.0011	-0.0007	0.0041	0.0408	0.0514	0.0669	0.0947	0.1682
		0.01%	1.1594	-0.0003	-0.0005	-0.0011	-0.0007	0.0041	0.0409	0.0516	0.0671	0.0946	0.1679
		0.1%	9.9994	-0.0007	-0.0006	-0.0013	-0.0007	0.0038	0.0431	0.0530	0.0677	0.0942	0.1654
	10-sec	0%	0.1581	-0.0009	-0.0012	0.0000	0.0008	0.0051	0.0403	0.0518	0.0682	0.0978	0.1744
		0.001%	0.2584	-0.0007	-0.0011	0.0001	0.0008	0.0052	0.0401	0.0516	0.0681	0.0978	0.1744
		0.01%	1.1597	-0.0006	-0.0011	0.0000	0.0007	0.0051	0.0408	0.0521	0.0683	0.0978	0.1740
		0.1%	9.9805	-0.0008	-0.0014	-0.0003	0.0003	0.0049	0.0462	0.0554	0.0701	0.0982	0.1717
	20-sec	0%	0.1582	-0.0012	-0.0007	0.0003	0.0009	0.0056	0.0347	0.0470	0.0646	0.0933	0.1691
		0.001%	0.2587	-0.0011	-0.0007	0.0003	0.0010	0.0056	0.0347	0.0469	0.0646	0.0934	0.1691
		0.01%	1.1609	-0.0005	-0.0004	0.0006	0.0013	0.0059	0.0353	0.0473	0.0648	0.0933	0.1687
		0.1%	9.9717	-0.0005	-0.0007	0.0003	0.0011	0.0056	0.0419	0.0520	0.0677	0.0940	0.1662

Table 4.6: ME and RMSE for NSR estimates M3 using $\overline{RV}_{KA^*}$ under the CTS and TTS schemes

Scheme	Sparsity	NSR	True NSR ($\times 10^4$)	ME ($\times 10^4$) 1-min	3-min	5-min	10-min	30-min	RMSE ($\times 10^4$) 1-min	3-min	5-min	10-min	30-min
CTS	5-sec	0%	0.1580	-0.0015	0.0010	0.0045	0.0608	0.1784	0.0408	0.2173	0.4671	1.3686	7.6554
		0.001%	0.2583	-0.0015	0.0008	0.0050	0.0600	0.1783	0.0409	0.2173	0.4670	1.3691	7.6551
		0.01%	1.1594	-0.0016	0.0002	0.0048	0.0599	0.1802	0.0416	0.2175	0.4665	1.3670	7.6438
		0.1%	9.9994	-0.0048	-0.0046	0.0026	0.0542	0.1772	0.0497	0.2198	0.4626	1.3475	7.5189
	10-sec	0%	0.1581	-0.0016	0.0031	0.0033	0.0599	0.1736	0.0413	0.2176	0.4689	1.3719	7.6538
		0.001%	0.2584	-0.0021	0.0030	0.0029	0.0603	0.1743	0.0412	0.2177	0.4690	1.3729	7.6512
		0.01%	1.1597	-0.0070	0.0024	0.0016	0.0580	0.1716	0.0430	0.2175	0.4684	1.3707	7.6339
		0.1%	9.9805	-0.0577	-0.0052	-0.0059	0.0471	0.1580	0.0804	0.2205	0.4664	1.3517	7.4858
	20-sec	0%	0.1582	-0.0170	0.0014	0.0016	0.0590	0.1732	0.0440	0.2166	0.4702	1.3663	7.6677
		0.001%	0.2587	-0.0269	0.0013	0.0021	0.0583	0.1737	0.0488	0.2166	0.4698	1.3653	7.6681
		0.01%	1.1609	-0.1160	0.0008	0.0014	0.0569	0.1711	0.1233	0.2174	0.4690	1.3620	7.6542
		0.1%	9.9717	-0.9879	-0.0178	-0.0132	0.0401	0.1500	0.9896	0.2283	0.4682	1.3373	7.5037
TTS	5-sec	0%	0.1580	0.0004	-0.0015	0.0038	0.0629	0.1904	0.0418	0.2178	0.4716	1.4060	8.3229
		0.001%	0.2583	0.0005	-0.0008	0.0054	0.0652	0.1995	0.0417	0.2179	0.4717	1.4066	8.3227
		0.01%	1.1594	0.0029	0.0060	0.0174	0.0895	0.2824	0.0420	0.2175	0.4710	1.4052	8.3114
		0.1%	9.9994	0.0252	0.0740	0.1344	0.3265	1.0854	0.0527	0.2290	0.4840	1.4172	8.2416
	10-sec	0%	0.1581	-0.0001	0.0071	0.0084	0.0514	0.2056	0.0450	0.2197	0.4811	1.3977	8.4093
		0.001%	0.2584	0.0005	0.0078	0.0096	0.0549	0.2152	0.0448	0.2197	0.4817	1.3973	8.4077
		0.01%	1.1597	0.0029	0.0146	0.0208	0.0793	0.2937	0.0458	0.2200	0.4819	1.3963	8.3900
		0.1%	9.9805	0.0270	0.0833	0.1369	0.3188	1.0849	0.0615	0.2354	0.4981	1.4055	8.2936
	20-sec	0%	0.1582	-0.0005	0.0085	0.0092	0.0460	0.2051	0.0459	0.2249	0.4647	1.4059	8.3888
		0.001%	0.2587	-0.0004	0.0095	0.0106	0.0486	0.2144	0.0460	0.2254	0.4646	1.4054	8.3887
		0.01%	1.1609	0.0021	0.0168	0.0232	0.0728	0.2947	0.0476	0.2264	0.4652	1.4029	8.3746
		0.1%	9.9717	0.0259	0.0853	0.1394	0.3088	1.0896	0.0695	0.2480	0.4838	1.4094	8.2728

Table 4.7: Empirical NSR estimates of 40 stocks

Stock	M1	M1A	M2	M2A	M3	Stock	M1	M1A	M2	M2A	M3
XOM	0.2972	0.2930	0.5029	0.4960	-0.0064	WMT	0.2355	0.2277	0.5263	0.5149	-0.1030
GE	0.5212	0.5152	0.4886	0.4802	0.6004	CVX	0.3698	0.3630	0.5871	0.5752	0.0130
IBM	0.2693	0.2564	0.3867	0.3613	-0.0020	JNJ	0.1845	0.1754	0.4032	0.3872	0.0036
T	0.2868	0.2840	0.3627	0.3674	0.2493	PG	0.6135	0.6030	0.8816	0.8675	-0.1380
JPM	0.0746	0.0716	0.1920	0.1874	-0.0848	WFC	0.0664	0.0614	0.1301	0.1228	-0.1325
KO	0.2998	0.2844	0.5412	0.5177	-0.0616	PFE	0.8523	0.8461	0.8729	0.8687	0.7280
C	6.6461	6.6515	6.6944	6.7040	6.4090	BAC	1.3697	1.3570	1.3009	1.2862	1.2926
SLB	0.2674	0.2627	0.6321	0.6234	-0.0776	MRK	0.1250	0.1148	0.3596	0.3467	-0.0834
PEP	0.1241	0.1115	0.4619	0.4405	-0.0808	VZ	0.3542	0.3438	0.4722	0.4566	0.0779
COP	0.0228	0.0193	0.2058	0.2022	-0.1118	GS	0.0852	0.0731	0.2469	0.2249	-0.1996
MCD	0.3643	0.3552	0.5448	0.5329	-0.0541	OXY	0.0442	0.0326	0.2834	0.2647	-0.0949
ABT	0.0630	0.0478	0.3865	0.3594	-0.0306	UTX	0.0095	-0.0062	0.1866	0.1599	-0.0745
UPS	0.1935	0.1808	0.3133	0.2965	-0.1725	F	0.8649	0.8334	0.8804	0.8490	0.7480
DIS	0.0863	0.0731	0.2690	0.2475	-0.1160	MMM	-0.0482	-0.0587	-0.2615	-0.2750	-0.4081
CAT	0.1926	0.1860	0.3940	0.3829	-0.1038	FCX	0.0844	0.0803	0.3373	0.3316	-0.1640
USB	0.1119	0.1034	0.1958	0.1846	-0.1161	MO	0.6505	0.6323	0.6575	0.6415	0.6197
AXP	0.1364	0.1247	0.3629	0.3443	-0.1685	BA	0.1011	0.0863	0.2494	0.2266	-0.0593
MDT	-0.1473	-0.1652	0.1268	0.0999	-0.1946	HD	0.1580	0.1480	0.2880	0.2714	-0.0663
CVS	0.1418	0.1266	0.3214	0.3060	-0.2244	EMC	0.2081	0.1979	0.2875	0.2760	0.0010
HAL	0.1610	0.1550	0.3816	0.3714	-0.2182	PNC	-0.0275	-0.0442	0.1436	0.1237	-0.2051

Notes: NSR is calculated under the TTS scheme. We use $\overline{RV}_{KA^*}$ at 1-min sampling frequency for M1, M1A and M3 and at 1-min versus 2-min sampling frequencies for M2 and M2A, 2010-2013. The figures are after multiplying by 10^4.

Table 4.8: Empirical NSR estimates of 40 stocks

Stock	M1	M1A	M2	M2A	M3	Stock	M1	M1A	M2	M2A	M3
XOM	0.0907	0.0892	0.5752	0.5692	0.3990	WMT	-0.0583	-0.0627	0.6827	0.6728	0.3820
GE	0.5462	0.5430	0.4562	0.4493	0.3183	CVX	0.1509	0.1496	0.7141	0.7040	0.5450
IBM	0.1533	0.1490	0.5312	0.5114	0.3096	JNJ	-0.0310	-0.0333	0.4323	0.4189	0.2086
T	0.2176	0.2076	0.5249	0.5253	0.2154	PG	0.3460	0.3392	1.2490	1.2362	0.8999
JPM	-0.0427	-0.0442	0.2606	0.2563	0.0802	WFC	0.0003	-0.0022	0.2682	0.2617	0.0476
KO	0.0561	0.0490	0.7099	0.6889	0.3749	PFE	0.8315	0.8250	0.9664	0.9595	0.7558
C	6.6024	6.6021	6.8326	6.8438	6.6471	BAC	1.4376	1.4340	1.3812	1.3670	1.2417
SLB	-0.0981	-0.0990	0.7098	0.7029	0.4719	MRK	-0.1073	-0.1138	0.4026	0.3890	0.1065
PEP	-0.2117	-0.2158	0.5357	0.5190	0.1754	VZ	0.2354	0.2299	0.5818	0.5684	0.3848
COP	-0.1564	-0.1598	0.2467	0.2419	0.0638	GS	-0.0712	-0.0739	0.3712	0.3540	0.1843
MCD	0.1830	0.1772	0.7008	0.6893	0.4434	OXY	-0.1975	-0.2018	0.3366	0.3204	0.1108
ABT	-0.2576	-0.2609	0.4102	0.3920	0.0259	UTX	-0.1623	-0.1687	0.2001	0.1765	-0.0408
UPS	0.0731	0.0642	0.4730	0.4547	0.2207	F	0.8486	0.8408	0.9296	0.8978	0.7625
DIS	-0.0922	-0.0971	0.3561	0.3387	0.0934	MMM	0.1462	0.1389	0.5025	0.4867	0.0652
CAT	-0.0088	-0.0109	0.4832	0.4742	0.3172	FCX	-0.1709	-0.1732	0.4380	0.4317	0.2816
USB	0.0278	0.0214	0.3044	0.2948	0.0622	MO	0.6420	0.6273	0.7331	0.7146	0.5607
AXP	-0.0958	-0.1004	0.4848	0.4699	0.2070	BA	-0.0389	-0.0464	0.3121	0.2912	0.0714
MDT	-0.4139	-0.4222	0.2299	0.2060	-0.1801	HD	0.0325	0.0288	0.4033	0.3906	0.1910
CVS	-0.0372	-0.0520	0.5898	0.5692	0.1723	EMC	0.1295	0.1208	0.4177	0.4039	0.1494
HAL	-0.0606	-0.0623	0.4530	0.4451	0.3109	PNC	-0.2050	-0.2152	0.2382	0.2186	-0.0751

Notes: NSR is calculated under the TTS scheme. We use $\overline{RV}_{KA^*}$ at 2-min sampling frequency for M1, M1A and M3 and at 1-min versus 3-min sampling frequencies for M2 and M2A, 2010-2013. The figures are after multiplying by 10^4.

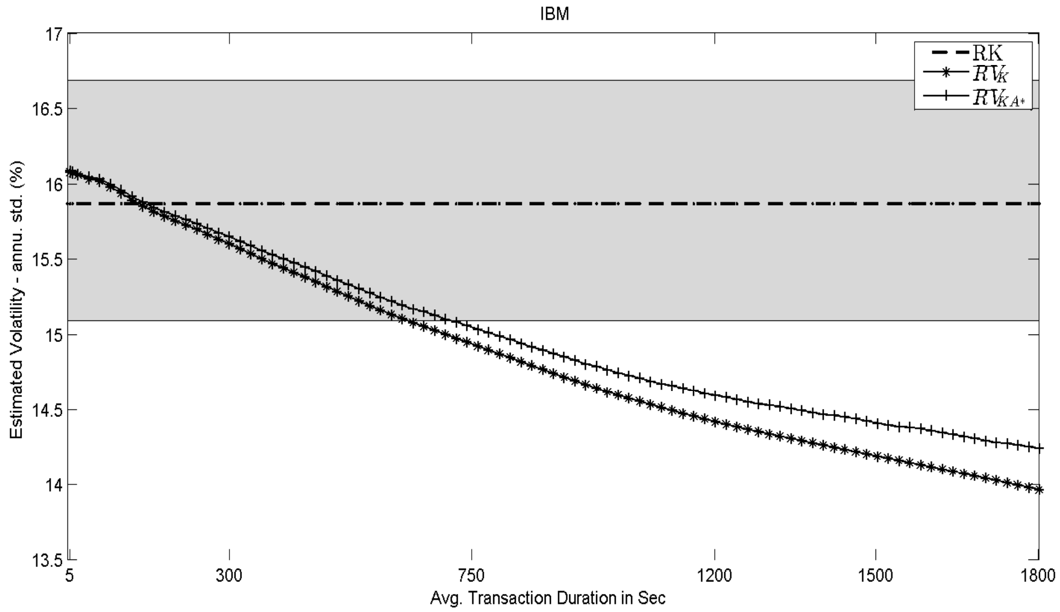

Figure 4.1: $\overline{RV}_K$ and $\overline{RV}_{KA^*}$ volatility signature plot for IBM under the TTS scheme, 2010-2013.

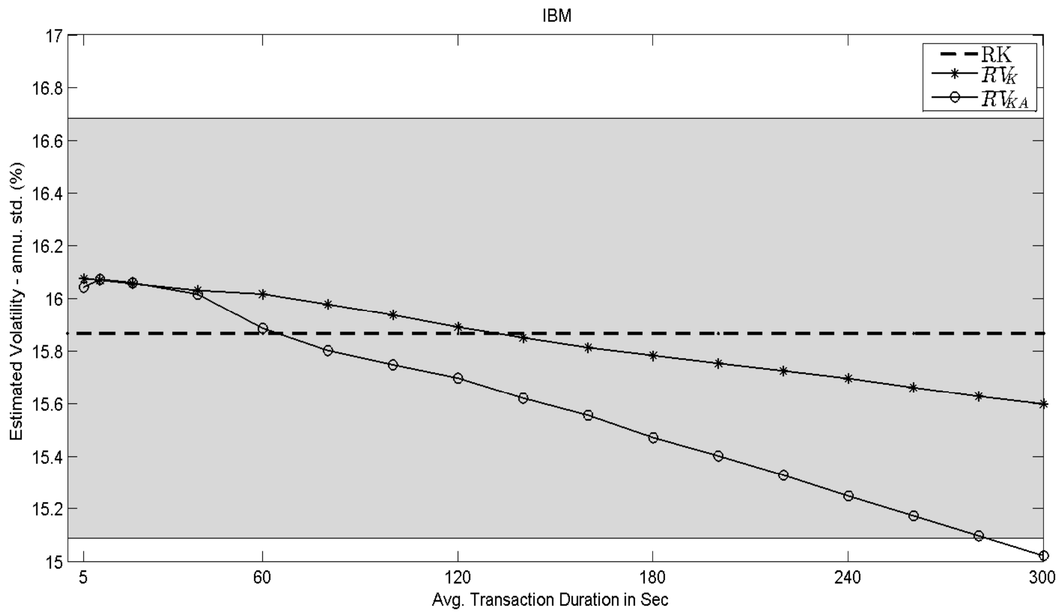

Figure 4.2: $\overline{RV}_K$ and $\overline{RV}_{KA}$ volatility signature plot for IBM under the TTS scheme, 2010-2013.

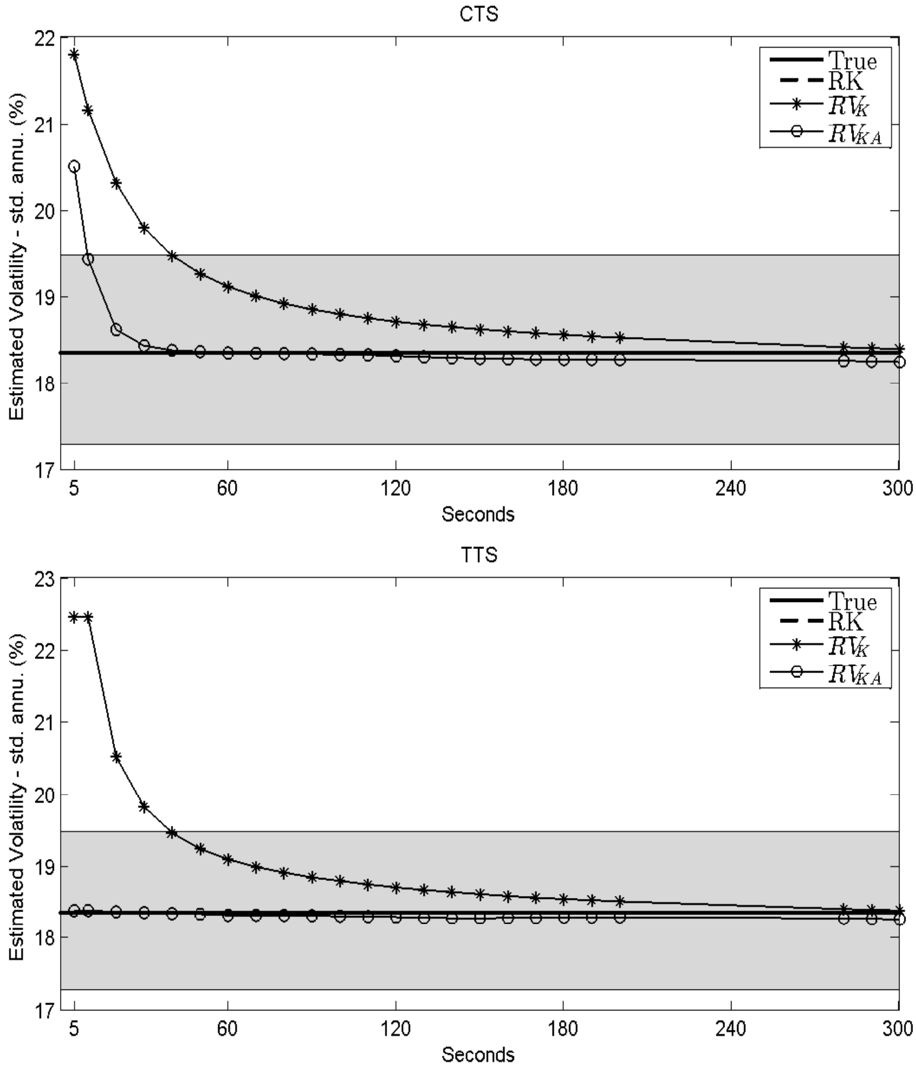

Figure 4.3: $\overline{RV}_K$ and $\overline{RV}_{KA}$ volatility signature plot of one simulation run, sparsity = 10-sec and NSR = 0.01%.

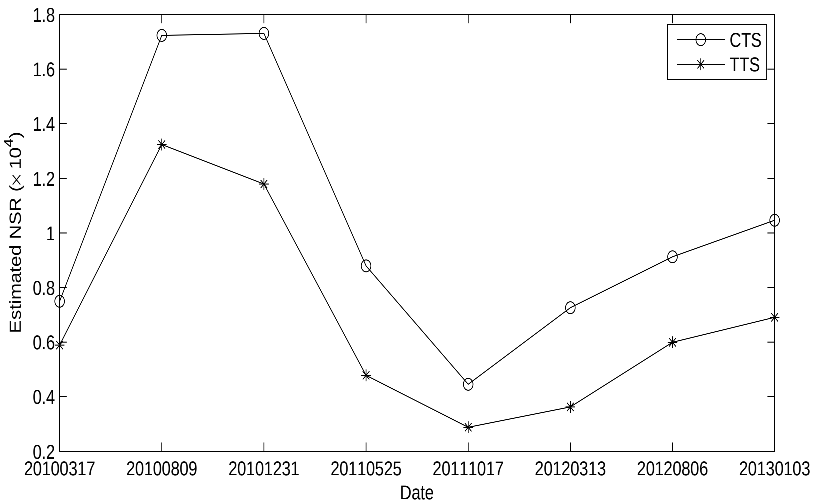

Figure 4.4: Average estimated NSR across 40 stocks over every 100 days, 2010-2013.

Chapter 5 Conclusion

My dissertation focuses on High-frequency financial data analysis using NYSE tick-by-tick transaction data. I investigate the intraday periodicity adjustment and its effect on intraday volatility estimation using the Time Transformation (TT) method in Chapter 2. In Chapter 3, I investigate the implementation and applications of the Business Time Sampling (BTS) scheme for high-frequency financial data. Specifically, I test the semi-martingale hypothesis using BTS returns and estimate the Integrated Volatility using either BTS returns or BTS durations. In Chapter 4, I investigate the market microstructure noise properties of typical NYSE stocks. I also propose two new microstructure noise-variance estimates in this chapter and estimate the Noise-to-signal Ratio (NSR) of NYSE stocks in period from 2010 to 2013.

In Chapter 2, I compare the Time Transformation (TT) method with the Duration Adjustment (DA) method. Advantages of the TT method over the DA method include its intuitive theoretical underpinning, its easier implementation, its easiness to switch between calendar time and diurnally adjusted time, among others. I then focus on adjusting the intraday periodicity in trading activity using the TT method and investigating effects of the adjustment on the estimation of intraday volatility. Our empirical results show that whether or not the duration data are adjusted for intraday periodicity has little impact on the daily volatility estimates using the ACD-ICV method. However, I document a clear regular intraday volatility smile, which shows a more prominent U-shape when the duration data are corrected for intraday periodicity than when it is not. I also show that multiple trades should be treated carefully in high-frequency financial data analysis by showing the different intraday volatility smiles when multiple trades are treated differently. Specifically, I show the intraday volatility smiles are more prominent when multiple trades are treated as separate trades.

Chapter 3 focuses on the intraday periodicity adjustment in volatility, which corresponds to the Business Time Sampling (BTS) scheme. I propose an easy-to-use time-transformation method to implement the BTS scheme and investigate applications of the BTS scheme. Using 40 typical NYSE stocks, I perform normality test to the jump-adjusted daily and weekly BTS returns. The computed results show that stock prices can be considered as discrete observations from a continuous-time jump-diffusion process. The BTS scheme performs better than the CTS and TTS schemes in yielding iid Gaussian returns, and it also performs better than the CTS and TTS schemes in estimating daily integrated volatility using the TRV method, with and without subsampling. I also show the superiority of the BTS durations over the price durations in estimating the daily integrated volatility using the Autoregressive Conditional Duration-Integrated Conditional Variance (ACD-ICV) method. I compare different Integrated Volatility estimates in the Monte Carlo simulation study, including the Tripower Realized Variance estimate, the Realized Kernel estimate, the original and modified ACD-ICV estimate. The modified ACD-ICV estimate, ME3, which models the high-frequency BTS durations using the ACD model, performs the best in reporting smaller RMSE values.

In Chapter 4, I investigate the noise properties of NYSE high-frequency transaction prices. I show the negative cross-correlation between the noise and stock returns. Selecting the sampling frequency is important to the microstructure noise-variance estimation. I then propose two noise-variance estimates M2 and M2A using difference of subsampling realized variance at two or multiple time scales. In the Monte Carlo simulation study, the new estimates performs better than other estimates by reporting lower ME and RMSE. Empirically the noise-to-signal ratio (NSR) are very small and 0.005% is a good approximate of the NSR for a typical NYSE stock in periods from 2010 to 2013.

Bibliography

AÏT-SAHALIA, Y. AND L. MANCINI (2008): "Out of sample forecasts of quadratic variation," *Journal of Econometrics*, 147, 17–33.

AÏT-SAHALIA, Y., P. A. MYKLAND, AND L. ZHANG (2005): "How often to sample a continuous-time process in the presence of market microstructure noise," *Review of Financial studies*, 18, 351–416.

AÏT-SAHALIA, Y., P. A. MYKLAND, AND L. ZHANG (2011): "Ultra high frequency volatility estimation with dependent microstructure noise," *Journal of Econometrics*, 160, 160–175.

ANDERSEN, T., T. BOLLERSELV, P. F. CHRISTOFFERSEN, AND F. X. DIEBOLD (2013): "Financial Risk Measurement for Financial Risk Management," *Handbook of the Economics of Finance, edited by G.Constantinides, M. Harris and R. Stulz*, 1127–1220.

ANDERSEN, T., T. BOLLERSLEV, F. DIEBOLD, AND P. LABYS (2001a): "Great realizations," *Risk*, 13, 105–108.

ANDERSEN, T. G. AND T. BOLLERSLEV (1997): "Intraday periodicity and volatility persistence in financial markets," *Journal of empirical finance*, 4, 115–158.

ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND H. EBENS (2001b): "The distribution of realized stock return volatility," *Journal of financial economics*, 61, 43–76.

ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2003): "Modeling and forecasting realized volatility," *Econometrica*, 71, 579–625.

ANDERSEN, T. G., T. BOLLERSLEV, AND D. DOBREV (2007): "No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and iid noise: Theory and testable distributional implications," *Journal of Econometrics*, 138, 125–180.

ANDERSEN, T. G., T. BOLLERSLEV, P. FREDERIKSEN, AND M. ØRREGAARD NIELSEN (2010): "Continuous-time models, realized volatilities, and testable distributional implications for daily stock returns," *Journal of Applied Econometrics*, 25, 233–261.

ANDERSEN, T. G., T. BOLLERSLEV, AND N. MEDDAHI (2011): "Realized volatility forecasting and market microstructure noise," *Journal of Econometrics*, 160, 220–234.

ANDERSEN, T. G., D. DOBREV, AND E. SCHAUMBURG (2012): "Jump-robust volatility estimation using nearest neighbor truncation," *Journal of Econometrics*, 169, 75–93.

ANDERSEN, T. G., D. DOBREV, AND E. SCHAUMBURG (2014): "A robust neighborhood truncation approach to estimation of integrated quarticity," *Econometric Theory*, 30, 3–59.

AWARTANI, B., V. CORRADI, AND W. DISTASO (2009): "Assessing market microstructure effects via realized volatility measures with an application to the dow jones industrial average stocks," *Journal of Business & Economic Statistics*, 27, 251–265.

BANDI, F. M. AND J. R. RUSSELL (2006): "Comment," *Journal of Business & Economic Statistics*, 24, 167–173.

BANDI, F. M. AND J. R. RUSSELL (2008): "Microstructure noise, realized variance, and optimal sampling," *The Review of Economic Studies*, 75, 339–369.

BANDI, F. M. AND J. R. RUSSELL (2011): "Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations," *Journal of Econometrics*, 160, 145–159.

BARNDORFF-NIELSEN, O. E. (2002): "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 253–280.

BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2008): "Designing realized kernels to measure the ex post variation of equity prices in the presence of noise," *Econometrica*, 76, 1481–1536.

BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2009): "Realized kernels in practice: Trades and quotes," *The Econometrics Journal*, 12, C1–C32.

BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2011): "Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading," *Journal of Econometrics*, 162, 149–169.

BARNDORFF-NIELSEN, O. E. AND N. SHEPHARD (2002): "Estimating quadratic variation using realized variance," *Journal of Applied econometrics*, 17, 457–477.

BARNDORFF-NIELSEN, O. E. AND N. SHEPHARD (2004): "Power and bipower variation with stochastic volatility and jumps," *Journal of financial econometrics*, 2, 1–37.

BARNDORFF-NIELSEN, O. E. AND N. SHEPHARD (2006): "Econometrics of testing for jumps in financial economics using bipower variation," *Journal of financial Econometrics*, 4, 1–30.

BARNDORFF-NIELSEN, O. E., N. SHEPHARD, AND M. WINKEL (2006): "Limit theorems for multipower variation in the presence of jumps," *Stochastic processes and their applications*, 116, 796–806.

BAUWENS, L. AND P. GIOT (2000): "The logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks," *Annales d'Economie et de Statistique*, 117–149.

BLACK, F. (1986): "Noise," *The journal of finance*, 41, 529–543.

CHRISTENSEN, K., R. OOMEN, AND M. PODOLSKIJ (2010): "Realised quantile-based estimation of the integrated variance," *Journal of Econometrics*, 159, 74–98.

DACOROGNA, M. M., U. A. MÜLLER, R. J. NAGLER, R. B. OLSEN, AND O. V. PICTET (1993): "A geographical model for the daily and weekly seasonal volatility in the foreign exchange market," *Journal of International Money and Finance*, 12, 413–438.

DAMBIS, K. E. (1965): "On the decomposition of continuous submartingales," *Theory of Probability & Its Applications*, 10, 401–410.

DIEBOLD, F. X. (2006): "On market microstructure noise and realized volatility," *Discussion of Hansen and Lunde*.

DIEBOLD, F. X. AND G. STRASSER (2013): "On the correlation structure of microstructure noise: A financial economic approach," *The Review of Economic Studies*, 80, 1304–1337.

DIONNE, G., P. DUCHESNE, AND M. PACURAR (2009): "Intraday Value at Risk (IVaR) using tick-by-tick data with application to the Toronto Stock Exchange," *Journal of Empirical Finance*, 16, 777–792.

DONG, Y. AND Y.-K. TSE (2014a): "Business Time Sampling Scheme and Its Application," Singpapore Management University, Working Paper.

DONG, Y. AND Y.-K. TSE (2014b): "A Study on Market Microstructure Noise," Singpapore Management University, Working Paper.

DUBINS, L. E. AND G. SCHWARZ (1965): "On continuous martingales," *Proceedings of the National Academy of Sciences of the United States of America*, 53, 913.

ENGLE, R. F. AND J. R. RUSSELL (1998): "Autoregressive conditional duration: a new model for irregularly spaced transaction data," *Econometrica*, 1127–1162.

FERNANDES, M. AND J. GRAMMIG (2006): "A family of autoregressive conditional duration models," *Journal of Econometrics*, 130, 1–23.

GHYSELS, E. AND A. SINKO (2011): "Volatility forecasting and microstructure noise," *Journal of Econometrics*, 160, 257–271.

GIOT, P. (2005): "Market risk models for intraday data," *The European Journal of Finance*, 11, 309–324.

GONÇALVES, S. AND N. MEDDAHI (2009): "Bootstrapping realized volatility," *Econometrica*, 77, 283–306.

HALL, P. AND C. C. HEYDE (1980): *Martingale limit theory and its application*, Academic press.

HANSEN, P. R. AND A. LUNDE (2006): "Realized variance and market microstructure noise," *Journal of Business & Economic Statistics*, 24, 127–161.

HASBROUCK, J. (1993): "Assessing the quality of a security market: A new approach to transaction-cost measurement," *Review of Financial Studies*, 6, 191–212.

HASBROUCK, J. (1999): "The dynamics of discrete bid and ask quotes," *The Journal of finance*, 54, 2109–2142.

HESTON, S. L. (1993): "A closed-form solution for options with stochastic volatility with applications to bond and currency options," *Review of financial studies*, 6, 327–343.

HUANG, X. AND G. TAUCHEN (2005): "The relative contribution of jumps to total price variance," *Journal of financial econometrics*, 3, 456–499.

JACOD, J., Y. LI, P. A. MYKLAND, M. PODOLSKIJ, AND M. VETTER (2009): "Microstructure noise in the continuous case: the pre-averaging approach," *Stochastic processes and their applications*, 119, 2249–2276.

JACOD, J., Y. LI, AND X. ZHENG (2013): "Statistical Properties of Microstructure Noise," *Available at SSRN 2212119*.

JACOD, J. AND A. N. SHIRYAEV (1987): *Limit theorems for stochastic processes*, vol. 1943877, Springer Berlin.

LIU, S. AND Y.-K. TSE (2015): "Intraday value-at-risk: an asymmetric autoregressive conditional duration approach," *Journal of Econometrics*, Forthcoming.

MYKLAND, P. A. (2012): "A Gaussian calculus for inference from high frequency data," *Annals of finance*, 8, 235–258.

NOWMAN, K. B. (1997): "Gaussian estimation of single-factor continuous time models of the term structure of interest rates," *The journal of Finance*, 52, 1695–1706.

O'HARA, M. (2015): "High frequency market microstructure," *Journal of Financial Economics*, forthcoming.

OOMEN, R. C. A. (2006a): "Comment," *Journal of Business & Economic Statistics*, 24, 195–202.

OOMEN, R. C. A. (2006b): "Properties of realized variance under alternative sampling schemes," *Journal of Business & Economic Statistics*, 24, 219–237.

PHILLIPS, P. C. B. AND J. YU (2005): "Comment on" Realized Variance and Market Microstructure Noise" by Peter R. Hansen and Asger Lunde," .

PHILLIPS, P. C. B. AND J. YU (2008): "Information loss in volatility measurement with flat price trading," .

PHILLIPS, P. C. B. AND J. YU (2011): "Corrigendum to A Gaussian approach for continuous time models of short-term interest rates(Yu, J. and PCB Phillips, Econometrics Journal, 4, 210–24)," *The Econometrics Journal*, 14, 126–129.

PODOLSKIJ, M., M. VETTER, ET AL. (2009): "Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps," *Bernoulli*, 15, 634–658.

TSE, Y.-K. AND Y. DONG (2014): "Intraday periodicity adjustments of transaction duration and their effects on high-frequency volatility estimation," *Journal of Empirical Finance*, 28, 352–361.

TSE, Y.-K. AND T. T. YANG (2012): "Estimation of high-frequency volatility: An autoregressive conditional duration approach," *Journal of Business & Economic Statistics*, 30, 533–545.

UBUKATA, M. AND K. OYA (2009): "Estimation and testing for dependence in market microstructure noise," *Journal of Financial Econometrics*, 7, 106–151.

WU, Z. (2012): "On the intraday periodicity duration adjustment of high-frequency data," *Journal of Empirical Finance*, 19, 282–291.

YU, J. AND P. C. B. PHILLIPS (2001): "A Gaussian approach for continuous time models of the short-term interest rate," *The Econometrics Journal*, 4, 210–224.

ZHANG, L. (2006): "Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach," *Bernoulli*, 12, 1019–1043.

ZHANG, L., P. A. MYKLAND, AND Y. AÏT-SAHALIA (2005): "A tale of two time scales," *Journal of the American Statistical Association*, 100.

ZHOU, B. (1996): "High-frequency data and volatility in foreign-exchange rates," *Journal of Business & Economic Statistics*, 14, 45–52.

ZHOU, B. (1998): "F-consistency, De-volatilization and normalization of high-frequency financial data," *Nonlinear Modelling of High Frequency Financial Time Series. Wiley, London*.

Appendix A Appendix of Chapter 2

A.1 Implementation of the Duration Adjustment Method

We divide each trading day (9:30-16:00) into 14 intervals, with the first and last intervals being 15 min each and all others being 30 min each. Let the number of trading days in the data set be T and the number of trades in the kth time interval on day i be N_{ik}, where $i = 1, \cdots, T$ and $k = 1, \cdots, 14$. For the purpose of computing N_{ik}, we may take multiple trades at the same time stamp as one trade or separate trades. Denote J_k as the number of seconds in the kth time interval. When multiple trades at the same time stamp are taken as one trade, we have $N_{ik} = \sum_{j=1}^{J_k} 1(N_{ikj>0})$, where $1(N_{ikj} > 0)$ is the indicator function taking value 1 when $N_{ikj} > 0$ and 0 otherwise, with N_{ikj} being the number of trades at the jth second in the kth interval on the ith trading day. However, when multiple trades at the same time stamp are taken as separate trades, we have $N_{ik} = \sum_{j=1}^{J_k} N_{ikj}$. For the kth time interval, the average number of trades over all trading days is computed as $N_k = (\sum_{i=1}^{T} N_{ik}/T)$. We then calculate the average trade duration in the kth time interval as $\bar{x}_k = J_k/N_k$. Let t_k be the mid-point of the kth time interval. We set the diurnal factor at the mid-point of each time interval to be equal to the standardized average trade duration of the interval so that $\phi(t_k) = \bar{x}_k / \left(\sum_{k=1}^{14} \bar{x}_k\right)$. Finally, the diurnal factor function $\phi(t)$ is calculated using the cubic spline smoothing with $\phi(t_k)$ as the knots, for $k = 1, \cdots, 14$.

A.2 Implementation of the Time Transformation Method

Let $t = 0, 1, \cdots, 23400$ denote the times 9:30:00, 9:30:01, $\cdots$, 16:00:00 in seconds, and let n_t denote the number of trades at time t aggregated over all days, for $t = 1, \cdots, 23400$. For the purpose of computing n_t, we may take multiple trades at the same time stamp as one trade or separate trades. We compute $N_t = \sum_{i=1}^{t} n_i$, for $t = 1, \cdots, 23400$. The time-transformation function $Q(t)$ is then computed as $Q(t) = N_t/N_{23400}$, for $t = 1, \cdots, 23400$, so that $\tilde{t} = 23400Q(t) = 23400N_t/N_{23400}$, where $\tilde{t}$ is the diurnally transformed time corresponding to calendar time t.

Appendix B Appendix of Chapter 3

B.1 Jump Detection Procedure

We remove price jumps using the sequential jump-adjustment procedure of Andersen *et al.* (2010) with some minor modifications. The test statistic Z_t for day t is computed as

$$Z_t = \sqrt{|\mathscr{H}_t|}\left[\frac{\ln V_{Rt} - \ln V_{Bt}}{((\mu_1^{-4} + 2\mu_1^{-2} - 5)Q_{Tt}BV_t^{-2})^{\frac{1}{2}}}\right], \tag{B.1.1}$$

where $\mu_1 = \sqrt{2/\pi}$, V_{Rt} is the realized volatility, V_{Bt} is the realized bipower volatility and Q_{Tt} is the realized tripower quarticity, and $\mathscr{H}_t$ is the grid on day t. The realized tripower quarticity Q_{Tt} is computed as

$$Q_{Tt} = \frac{1}{|\mathscr{H}_t|}\left[\mu_{\frac{4}{3}}^{-3}\sum_{j=2}^{|\mathscr{H}_t|-1}|r_{i,-}|^{\frac{4}{3}}|r_i|^{\frac{4}{3}}|r_{i,+}|^{\frac{4}{3}}\right], \tag{B.1.2}$$

where $r_i = Y_{i,+} - Y_i$ for $Y_{i,+}, Y_i \in \mathscr{H}_t$ and $\mu_{\frac{4}{3}} = 2^{\frac{2}{3}}\Gamma(7/6)/\Gamma(1/2)$. Under the condition of no jumps, Z_t is approximately standard normal. At the significance level α, a jump is considered to be significant if $Z_t > z_{1-\alpha}$, where z_α is the α-quantile of the standard normal distribution. When $Z_t > z_{1-\alpha}$, we delete the return that has the maximum absolute value among all returns and repeat the jump test again for the remaining returns. This recursive procedure stops when $Z_t \leq z_{1-\alpha}$.

B.2 Computation of the BT Transformation Function

To compute the BT transformation function $Q(t_k) = N_{t_k}/N_{t_K}$, where $N_{tk} = \sum_{i=1}^{k} V_i$ for $k = 1, \cdots, K$, we need to calculate the intraday integrated volatility V_i. To alleviate the effects of market microstructure noise and price jumps we adopt the Tripower Realized Volatility (TRV) method of Barndorff-Nielsen *et al.* (2006). Suppose there are S subsampling grids and the sth subgrid is denoted by $\mathscr{H}_s$. Under subgrid $\mathscr{H}_s$, for trading day d, denote its consecutive returns by $\{r_{d,1}^{(s)}, r_{d,2}^{(s)}, r_{d,3}^{(s)}, \cdots, r_{d,n_s}^{(s)}\}$. We define

$$V_{Td,i}^{(s)} = \xi_{\frac{2}{3}}^{-3} |r_{d,i-1}^{(s)}|^{\frac{2}{3}} |r_{d,i}^{(s)}|^{\frac{2}{3}} |r_{d,i+1}^{(s)}|^{\frac{2}{3}},$$

for $i = 2, \cdots, n_s - 1$, where $\xi_k = 2^{\frac{k}{2}} \Gamma((k+1)/2)/\Gamma(1/2)$ for $k > 0$ and $\Gamma(\cdot)$ is the gamma function. Moreover, $V_{Td,1}^{(s)} = V_{Td,2}^{(s)}$ and $V_{Td,n_s}^{(s)} = V_{Td,n_s-1}^{(s)}$. Since the estimate $V_{Td,i}^{(s)}$ may have large fluctuations, we group every g $V_{Td,i}^{(s)}$ together to obtain the estimated intraday integrated volatility at 10-min average sampling frequency. Thus, the pth intraday integrated volatility on day d under subgrid $\mathscr{H}_s$ is $V_{d,p}^{(s)} = \sum_{i=p(g-1)+1}^{pg} V_{Td,i}^{(s)}$. The corresponding end point of $V_{d,p}^{(s)}$ is the end point of $r_{d,pg}^{(s)}$ and is denoted by $t_{d,p}^{(s)}$. Finally, the estimated pth intraday integrated volatility $V_{d,p}$ and its corresponding end point $t_{d,p}$ can be obtained by taking the average over all subsamples. That is, $V_{d,p} = \frac{1}{S}\sum_{s=1}^{S} V_{d,p}^{(s)}$ and $t_{d,p} = \frac{1}{S}\sum_{s=1}^{S} t_{d,p}^{(s)}$. $V_{d,p}$ and $t_{d,p}$ pooled over all m trading days form V_K for $k = 1, \cdots, K$ and $\mathscr{H}_V$, respectively, for the computation of the time-transformation function $Q(t)$. The number of subsamples S is selected to obtain subgrids of approximately 1-min sampling frequency.

B.3 Supplementary Material

B.3.1 Intraday Periodicity and the BTS Scheme

We examine the intraday periodicity in volatility and trading activity of the top 40 stocks (by market capitalization as of 2010) from the NYSE from January 2010 to April 2013. We calculate the squared 1-min calendar-time returns on each trading day and take the average at each 1-min time interval over all trading days in the sample period. These series are treated as 1-min intraday realized volatility estimates, which are plotted on the left-hand column of Figure B.1 (realized volatility is reported in annualized standard deviation in percent). We also calculate the total number of transactions at each sec from 09:30 to 16:00 over all trading days, which are plotted on the right-hand column of Figure B.1. The intraday periodicity in volatility are quite different from that in trading activity, which shows that we cannot use the TTS scheme to approximate the BTS scheme.

In Table B.1 we report the proportion of detected jumps at different sampling frequencies under different sampling schemes using the sequential jump-detection procedure of Andersen et al. (2010). We investigate cases when the sampling frequency is equal to 1 min, 5 min or 10 min. We also calculate the proportion of trading days for which we reject the normality assumption at the 5% significance level for the CTS, TTS and BTS returns at different sampling frequencies with jump adjustment. Table B.2 presents the results for the returns with jumps deleted using the sequential method of Andersen *et al.* (2010) at the significance level of 1%. Table B.3 reports the results for the returns without jump adjustment.

B.3.2 Simulation Results for the Integrated Volatility Estimates

Tables B.4-B.9 provide the mean error (ME) and root mean-squared error (RMSE) of the daily integrated volatility estimates using the TRV method with and without subsampling under different sampling schemes. Table B.9 is the same as Table 4 of the paper. Tables B.10-B.12 report the ME and RMSE of the RK and the ACD-ICV estimates for MD2, MD3 and MD4.

Table B.1: Proportion of detected jumps

	1-min			5-min			10-min		
Stock	CTS	TTS	BTS	CTS	TTS	BTS	CTS	TTS	BTS
XOM	4.70	2.55	1.91	4.67	3.90	2.77	7.36	5.84	4.73
WMT	18.10	8.79	7.35	5.78	3.81	4.38	8.16	5.87	5.36
GE	20.54	21.15	26.07	10.58	8.41	5.95	10.82	8.23	6.42
CVX	3.49	1.68	1.93	4.41	3.13	3.07	6.64	5.94	4.86
IBM	5.99	1.80	2.71	4.66	3.46	3.75	6.51	5.42	4.62
JNJ	16.62	9.88	4.78	5.55	4.57	4.00	7.59	6.42	5.41
T	21.87	20.54	25.55	9.13	7.68	5.12	9.77	6.82	6.60
PG	17.33	9.33	4.99	5.82	4.54	4.18	7.95	6.06	5.33
JPM	8.50	3.66	3.57	4.92	3.23	3.22	7.50	5.33	4.70
WFC	17.96	10.43	6.87	6.06	4.65	3.47	7.68	5.61	4.73
KO	17.93	11.03	7.46	6.70	4.21	4.18	7.79	6.06	5.46
PFE	19.62	20.48	26.02	12.96	11.13	9.28	10.02	9.09	6.90
C	8.50	6.15	7.96	8.92	8.60	8.44	10.56	10.05	8.60
BAC	21.83	22.43	22.01	12.91	11.50	12.97	12.59	8.41	6.82
SLB	2.75	1.44	1.91	4.59	3.27	3.13	6.68	5.66	4.41
MRK	22.29	15.45	11.96	6.44	5.44	4.24	8.50	6.17	5.78
PEP	18.00	8.97	5.72	6.15	4.71	4.07	7.03	5.98	5.19
VZ	23.07	17.65	17.31	7.50	5.90	4.86	8.64	6.97	5.80
COP	8.27	4.38	2.83	5.04	3.46	3.26	7.32	5.21	4.40
GS	6.50	1.49	2.86	4.93	3.81	3.34	7.56	5.76	5.05
MCD	10.11	3.94	4.07	5.13	3.80	3.89	7.17	5.89	5.00
OXY	5.93	1.92	2.58	4.78	3.78	3.29	6.94	5.52	4.30
ABT	19.87	12.64	8.45	5.97	4.62	4.05	7.78	6.21	5.68
UTX	10.33	3.12	3.39	5.44	3.41	3.48	7.57	5.23	5.46
UPS	12.64	3.64	3.41	5.35	4.34	3.79	6.63	5.73	5.02
F	19.73	20.02	27.25	13.73	10.46	7.08	11.92	8.82	6.80
DIS	17.89	9.75	7.61	5.19	4.44	4.51	7.62	6.15	5.57
MMM	9.17	2.66	2.89	4.99	3.68	3.50	7.39	6.45	5.06
CAT	3.62	1.65	2.25	4.74	3.40	3.43	6.79	5.06	4.35
FCX	5.90	2.46	2.63	4.21	2.94	2.99	6.62	5.29	4.55
USB	21.82	15.96	13.08	7.95	4.96	4.32	8.39	5.83	4.89
MO	20.09	20.12	26.06	11.25	9.36	7.33	10.90	7.76	7.14
AXP	12.20	4.24	3.61	4.45	3.55	3.39	7.90	5.30	4.33
BA	10.05	2.72	3.14	4.59	4.24	3.54	7.64	6.03	5.12
MDT	23.52	12.86	10.32	6.67	5.25	4.10	7.60	6.23	5.87
HD	17.39	9.08	8.20	5.40	4.59	4.78	7.58	5.99	5.72
CVS	24.02	13.18	12.58	6.64	4.29	5.09	7.47	6.38	6.26
EMC	22.96	18.22	20.48	7.57	6.19	4.86	8.81	6.60	5.62
HAL	8.03	2.78	3.41	4.66	3.75	3.30	6.73	5.96	5.00
PNC	13.63	3.21	3.14	5.28	3.64	3.34	7.43	5.40	5.23

Notes: The figures in the table are the proportions of detected jumps in percentage, 2010-2013.

Table B.2: Rejection proportion of the normality hypothesis for no-jump returns under different sampling schemes

	1-min			5-min			10-min		
Stock	CTS	TTS	BTS	CTS	TTS	BTS	CTS	TTS	BTS
XOM	93.86	86.04	81.95	18.41	12.64	8.06	7.46	6.14	4.57
WMT	99.76	99.76	98.44	43.08	25.27	17.45	13.36	7.34	5.90
GE	100.00	100.00	99.88	82.55	72.80	67.51	35.42	29.72	21.18
CVX	89.33	78.19	68.23	13.36	7.62	5.86	6.46	5.86	3.52
IBM	83.49	76.39	41.93	13.25	12.65	4.82	7.95	4.34	5.42
JNJ	99.40	98.80	97.11	39.16	29.28	15.66	14.10	8.55	6.39
T	100.00	100.00	99.88	75.93	67.39	53.31	30.45	24.19	17.81
PG	100.00	98.80	97.11	36.58	25.27	14.92	11.43	8.67	5.54
JPM	99.04	95.43	92.18	27.08	16.61	8.42	10.35	6.74	6.14
WFC	99.88	99.18	97.54	45.02	31.18	21.57	15.24	10.43	6.57
KO	99.53	98.48	97.30	42.32	32.59	18.05	13.83	9.50	7.62
PFE	100.00	100.00	100.00	87.22	82.65	71.51	39.39	35.52	26.73
C	99.65	97.89	94.61	57.21	47.71	45.37	46.89	42.56	39.79
BAC	100.00	100.00	99.88	83.35	77.37	69.75	47.48	37.63	30.25
SLB	85.09	75.70	48.71	15.73	11.03	5.52	6.81	5.99	3.17
MRK	100.00	99.65	98.94	55.10	48.30	30.13	18.52	13.48	10.32
PEP	99.53	98.36	96.72	37.98	29.43	13.13	11.72	9.73	3.87
VZ	100.00	100.00	99.88	60.96	50.06	32.24	23.21	17.35	10.67
COP	96.13	91.91	84.64	22.30	16.65	12.19	9.61	6.57	4.57
GS	82.53	68.00	24.15	13.60	11.96	4.92	7.98	5.74	2.70
MCD	98.59	95.90	92.73	27.67	20.16	9.38	11.37	9.50	5.75
OXY	85.93	83.70	51.35	15.36	13.72	5.74	8.09	6.45	4.22
ABT	99.77	99.41	98.71	47.13	36.34	21.22	16.30	13.95	6.57
UTX	97.30	94.49	84.41	25.56	20.28	7.27	11.61	8.80	5.16
UPS	97.66	95.08	90.04	24.03	19.58	8.21	12.19	9.26	3.40
F	100.00	100.00	100.00	87.22	78.90	73.97	43.85	34.11	27.67
DIS	99.77	99.53	98.59	41.85	30.13	18.52	13.13	8.91	7.39
MMM	96.48	91.44	79.72	21.57	15.94	7.15	6.69	7.85	4.69
CAT	89.21	75.73	49.71	17.00	10.67	5.04	6.92	6.21	3.05
FCX	89.33	74.44	59.55	16.30	10.90	7.62	8.80	5.16	4.81
USB	100.00	99.77	99.41	55.69	45.13	29.66	20.87	14.07	9.14
MO	100.00	100.00	99.88	86.64	77.96	64.13	39.39	30.60	23.24
AXP	98.71	96.37	92.26	25.21	18.76	11.84	10.08	6.68	5.04
BA	97.77	93.67	81.83	26.73	20.28	6.10	12.78	9.15	3.52
MDT	100.00	100.00	98.94	48.71	36.74	19.72	15.75	12.56	6.92
HD	100.00	99.65	98.36	41.15	32.24	18.29	13.83	10.67	7.16
CVS	99.88	99.53	99.06	49.41	35.33	18.78	13.73	12.09	8.22
EMC	100.00	100.00	99.77	65.65	54.40	35.29	22.51	18.41	9.96
HAL	96.37	92.03	85.58	22.51	14.42	8.44	8.56	6.21	3.63
PNC	98.12	93.20	83.24	24.74	17.82	6.45	9.73	8.22	5.75

Notes: The normality test is implemented for the high-frequency returns (2010-2013) without jumps and the significance level is 5%. The figures are in percentage of proportion of the trading days that reject the normality hypothesis. Total number of trading days of each stock investigated in our paper ranges from 830 to 853.

Table B.3: Rejection proportion of the normality hypothesis for returns under different sampling schemes

	1-min			5-min			10-min		
Stock	CTS	TTS	BTS	CTS	TTS	BTS	CTS	TTS	BTS
XOM	94.22	86.76	81.35	19.37	13.48	10.35	7.58	7.82	5.05
WMT	99.76	99.64	98.19	41.16	25.99	20.34	13.24	8.78	6.26
GE	100.00	100.00	99.88	81.95	73.16	68.83	34.54	28.52	22.86
CVX	88.86	76.79	68.93	15.24	8.21	7.62	7.74	6.80	4.69
IBM	85.54	77.59	43.01	14.34	13.01	7.23	9.04	4.70	6.14
JNJ	99.64	98.80	97.11	35.66	29.52	18.19	15.18	9.88	7.23
T	100.00	100.00	99.88	75.69	67.51	53.91	31.17	24.79	17.21
PG	100.00	98.68	96.99	36.46	25.27	16.49	12.64	10.47	7.22
JPM	99.04	95.67	92.78	27.20	17.45	9.99	11.31	7.22	6.86
WFC	99.88	99.30	97.30	43.26	32.36	23.56	14.54	11.84	8.21
KO	99.30	98.48	97.30	42.91	32.12	20.28	15.01	9.96	9.61
PFE	100.00	100.00	100.00	85.93	81.48	72.10	39.16	35.05	27.67
C	99.65	97.89	94.61	57.91	47.36	46.66	47.01	43.26	40.33
BAC	100.00	100.00	99.88	83.35	77.61	70.11	47.13	36.46	31.77
SLB	84.51	74.77	49.88	14.91	11.85	6.22	8.22	6.10	4.23
MRK	100.00	99.65	98.83	54.87	48.30	31.07	19.11	14.42	11.02
PEP	99.53	98.36	96.37	38.69	30.83	14.42	12.08	10.43	4.57
VZ	100.00	100.00	99.77	61.55	49.59	33.65	24.74	17.47	11.72
COP	96.01	91.79	85.23	22.63	16.88	14.07	10.20	6.45	5.63
GS	82.06	68.00	25.32	14.89	12.08	6.10	8.91	6.57	3.40
MCD	98.71	95.66	92.61	27.08	20.75	11.14	13.01	10.32	6.33
OXY	86.28	83.59	50.53	16.41	14.42	7.03	8.68	7.15	4.81
ABT	99.77	99.53	98.24	45.25	34.94	23.09	15.94	13.60	6.33
UTX	96.95	94.72	84.64	26.03	19.93	8.91	12.31	9.50	6.21
UPS	97.54	94.37	90.15	25.79	21.10	8.91	12.43	9.14	3.75
F	100.00	100.00	100.00	86.75	78.90	75.03	42.09	34.82	28.02
DIS	99.77	99.53	98.83	40.56	30.95	21.92	13.13	9.96	7.97
MMM	96.25	90.74	80.66	22.86	17.23	8.44	8.44	9.14	5.28
CAT	89.45	74.79	51.00	18.87	11.72	5.74	8.79	6.45	3.87
FCX	89.10	74.33	60.02	17.82	11.37	8.44	9.03	4.92	5.63
USB	100.00	99.77	99.30	54.63	44.20	30.60	20.63	14.30	10.79
MO	100.00	100.00	100.00	86.75	76.44	64.48	38.10	29.54	23.56
AXP	98.83	96.37	92.61	26.85	18.64	13.48	11.49	7.97	6.45
BA	97.42	93.32	81.59	26.38	20.98	7.85	12.78	8.91	4.92
MDT	100.00	100.00	99.18	48.83	35.80	20.89	16.08	12.68	7.75
HD	100.00	99.53	98.71	41.74	29.54	20.16	15.36	10.20	7.97
CVS	100.00	99.41	99.06	48.00	33.80	21.01	14.67	11.38	7.98
EMC	100.00	100.00	99.88	64.95	53.81	37.05	21.10	18.41	10.67
HAL	96.01	91.21	85.23	23.56	14.77	10.32	9.61	7.74	4.92
PNC	97.89	93.08	82.30	25.09	16.53	7.85	11.14	9.26	7.85

Notes: The normality test is implemented for the high-frequency returns (2010-2013) without jump adjustment and the significance level is 5%. The figures are in percentage of proportion of the trading days that reject the normality hypothesis. Total number of trading days of each stock investigated in our paper ranges from 830 to 853.

Table B.4: ME and RMSE of daily volatility estimates using the TRV method without subsampling under the CTS scheme

Sparsity	NSR	Model	ME					RMSE				
			1-min	2-min	3-min	5-min	10-min	1-min	2-min	3-min	5-min	10-min
5-sec	0.005%	MD1	0.2765	0.0018	-0.1603	-0.4215	-0.9202	1.8804	2.6007	3.1602	4.0590	5.8376
		MD2	0.0356	-0.0846	-0.1627	-0.3037	-0.6353	1.2804	1.7714	2.1589	2.7816	3.9611
		MD3	0.3045	0.0713	-0.0585	-0.2489	-0.7224	1.7884	2.4828	3.0502	3.9652	5.6334
		MD4	0.4156	0.2356	0.1679	0.0247	-0.4923	1.9250	2.7166	3.3612	4.2869	5.7800
		MD5	0.7956	0.7680	0.8243	0.9138	0.9625	2.1969	2.9971	3.6751	4.7167	6.7030
	0.01%	MD1	1.0475	0.3853	0.0949	-0.2659	-0.8453	2.1758	2.6506	3.1766	4.0599	5.8367
		MD2	0.5479	0.1720	0.0076	-0.2041	-0.5874	1.4343	1.7973	2.1677	2.7840	3.9615
		MD3	1.0395	0.4387	0.1845	-0.1021	-0.6511	2.0729	2.5383	3.0736	3.9702	5.6359
		MD4	1.1810	0.6181	0.4212	0.1775	-0.4130	2.2297	2.7762	3.3845	4.2916	5.7746
		MD5	1.5738	1.1545	1.0821	1.0711	1.0411	2.6187	3.1405	3.7595	4.7610	6.7240
	0.02%	MD1	2.5583	1.1484	0.6060	0.0384	-0.6950	3.2620	2.9186	3.2798	4.0837	5.8387
		MD2	1.5586	0.6783	0.3466	-0.0004	-0.4867	2.1291	1.9585	2.2235	2.7988	3.9643
		MD3	2.4829	1.1694	0.6695	0.1853	-0.5064	3.1055	2.7979	3.1751	4.0007	5.6464
		MD4	2.6748	1.3745	0.9285	0.4797	-0.2590	3.3024	3.0429	3.4905	4.3217	5.7700
		MD5	3.0965	1.9223	1.5958	1.3815	1.1966	3.8041	3.5453	3.9743	4.8697	6.7727
10-sec	0.005%	MD1	-0.2079	-0.1010	-0.1970	-0.4394	-0.9293	1.8557	2.5940	3.1417	4.0793	5.8419
		MD2	-0.2901	-0.1496	-0.1914	-0.3297	-0.6460	1.3079	1.7686	2.1641	2.7849	3.9646
		MD3	-0.1525	-0.0152	-0.1067	-0.2822	-0.7214	1.7672	2.4800	3.0529	3.9676	5.6270
		MD4	0.0570	0.1687	0.1243	-0.0020	-0.5132	1.8638	2.7000	3.3345	4.2761	5.7746
		MD5	0.3063	0.6868	0.7751	0.8785	0.9542	2.0505	2.9700	3.6640	4.7304	6.6935
	0.01%	MD1	0.5875	0.2843	0.0584	-0.2834	-0.8549	1.9722	2.6322	3.1550	4.0786	5.8397
		MD2	0.2392	0.1065	-0.0200	-0.2281	-0.5947	1.3340	1.7817	2.1696	2.7857	3.9646
		MD3	0.5998	0.3509	0.1382	-0.1391	-0.6520	1.8899	2.5251	3.0705	3.9718	5.6282
		MD4	0.8420	0.5530	0.3749	0.1514	-0.4360	2.0561	2.7543	3.3594	4.2797	5.7719
		MD5	1.1028	1.0782	1.0365	1.0340	1.0280	2.3484	3.1071	3.7484	4.7752	6.7165
	0.02%	MD1	2.1353	1.0514	0.5703	0.0207	-0.7064	2.9200	2.8695	3.2476	4.0969	5.8447
		MD2	1.2704	0.6162	0.3200	-0.0283	-0.4953	1.9108	1.9282	2.2218	2.7940	3.9630
		MD3	2.0703	1.0814	0.6230	0.1480	-0.5064	2.7838	2.7605	3.1571	3.9975	5.6322
		MD4	2.3616	1.3139	0.8782	0.4572	-0.2889	3.0457	3.0102	3.4572	4.3033	5.7685
		MD5	2.6595	1.8532	1.5518	1.3425	1.1787	3.4399	3.5026	3.9594	4.8790	6.7581
20-sec	0.005%	MD1	-3.3631	-0.6075	-0.4036	-0.4887	-0.9360	3.8969	2.6448	3.1505	4.0686	5.8363
		MD2	-2.3761	-0.4721	-0.3109	-0.3650	-0.6672	2.7582	1.8149	2.1613	2.7968	3.9626
		MD3	-3.0192	-0.4931	-0.2803	-0.3365	-0.7559	3.5111	2.5065	3.0455	3.9633	5.6222
		MD4	-2.2453	-0.2437	-0.0441	-0.0970	-0.5295	2.9432	2.6686	3.2973	4.2452	5.7747
		MD5	-2.9017	0.1677	0.5882	0.8210	0.9329	3.6002	2.8912	3.6233	4.7002	6.6668
	0.01%	MD1	-2.5819	-0.2065	-0.1414	-0.3367	-0.8616	3.2493	2.6054	3.1512	4.0667	5.8367
		MD2	-1.8555	-0.2040	-0.1397	-0.2630	-0.6202	2.3096	1.7829	2.1573	2.7960	3.9652
		MD3	-2.2787	-0.1109	-0.0339	-0.1870	-0.6846	2.9148	2.4815	3.0467	3.9625	5.6224
		MD4	-1.4825	0.1513	0.2138	0.0542	-0.4511	2.4112	2.6634	3.3044	4.2466	5.7679
		MD5	-2.1231	0.5701	0.8544	0.9801	1.0126	3.0148	2.9675	3.6960	4.7462	6.6840
	0.02%	MD1	-1.0765	0.5809	0.3751	-0.0345	-0.7144	2.2851	2.7121	3.2132	4.0855	5.8370
		MD2	-0.8510	0.3208	0.2058	-0.0558	-0.5176	1.6202	1.8455	2.1936	2.8055	3.9647
		MD3	-0.8468	0.6410	0.4596	0.1059	-0.5399	2.0605	2.6001	3.1160	3.9837	5.6253
		MD4	-0.0140	0.9322	0.7264	0.3574	-0.2999	1.9210	2.8273	3.3794	4.2665	5.7655
		MD5	-0.6113	1.3614	1.3800	1.2937	1.1634	2.2695	3.2667	3.8906	4.8464	6.7261

Notes: ME and RMSE are of the annualized standard deviation in percentage. The average true daily integrated volatility is around 40% for MD1, 27% for MD2, 36% for MD3, 28% for MD4 and 40% for MD5. MD1 and MD2 are the Heston model at different volatility level. MD3 is the two-factor stochastic volatility model with intraday volatility periodicity. MD4 is the deterministic volatility model with intraday volatility periodicity and MD5 is the Heston model (MD1) with price jumps. The first column indicates the average duration of the observed simulated transactions.

Table B.5: ME and RMSE of daily volatility estimates using the TRV method without subsampling under the TTS scheme

Sparsity	NSR	Model	ME 1-min	ME 2-min	ME 3-min	ME 5-min	ME 10-min	RMSE 1-min	RMSE 2-min	RMSE 3-min	RMSE 5-min	RMSE 10-min
5-sec	0.005%	MD1	-0.0486	-0.1645	-0.2494	-0.3872	-0.7403	1.8625	2.5866	3.1463	4.0825	5.7765
		MD2	-0.1811	-0.1996	-0.2175	-0.2841	-0.4918	1.2959	1.7819	2.1579	2.7673	3.9429
		MD3	0.0039	-0.0570	-0.1106	-0.1906	-0.5129	1.7806	2.4898	3.0538	3.9685	5.6190
		MD4	0.1888	0.1297	0.1188	0.0329	-0.4067	1.9010	2.7056	3.3554	4.2846	5.7511
		MD5	0.4782	0.6140	0.7428	0.9141	1.1668	2.1260	2.9674	3.6460	4.7036	6.7239
	0.01%	MD1	0.7339	0.2245	0.0067	-0.2307	-0.6617	2.0412	2.6161	3.1551	4.0850	5.7784
		MD2	0.3416	0.0613	-0.0456	-0.1785	-0.4425	1.3667	1.7910	2.1604	2.7716	3.9465
		MD3	0.7511	0.3136	0.1357	-0.0418	-0.4377	1.9593	2.5286	3.0682	3.9741	5.6182
		MD4	0.9657	0.5185	0.3762	0.1855	-0.3297	2.1335	2.7575	3.3737	4.2903	5.7488
		MD5	1.2671	1.0047	1.0037	1.0713	1.2440	2.4692	3.0959	3.7283	4.7519	6.7487
	0.02%	MD1	2.2705	0.9999	0.5199	0.0797	-0.5061	3.0350	2.8399	3.2403	4.1106	5.7844
		MD2	1.3618	0.5754	0.2949	0.0255	-0.3415	1.9848	1.9264	2.2147	2.7907	3.9514
		MD3	2.2102	1.0523	0.6249	0.2517	-0.2942	2.9013	2.7589	3.1564	4.0050	5.6265
		MD4	2.4765	1.2806	0.8859	0.4938	-0.1747	3.1497	3.0070	3.4744	4.3188	5.7457
		MD5	2.8118	1.7841	1.5249	1.3856	1.4003	3.5912	3.4727	3.9386	4.8604	6.8042
10-sec	0.005%	MD1	-0.5513	-0.4358	-0.4229	-0.4977	-0.7675	1.9652	2.6115	3.1903	4.0962	5.7850
		MD2	-0.5171	-0.3709	-0.3367	-0.3680	-0.5283	1.3995	1.8032	2.1732	2.7816	3.9376
		MD3	-0.4485	-0.3124	-0.2851	-0.3108	-0.5679	1.8592	2.4968	3.0428	3.9461	5.6049
		MD4	-0.1800	-0.0730	-0.0693	-0.0616	-0.4648	1.9217	2.6951	3.3350	4.2743	5.7368
		MD5	-0.0295	0.3201	0.5488	0.8034	1.0889	2.0870	2.9066	3.6022	4.7083	6.6976
	0.01%	MD1	0.2402	-0.0316	-0.1606	-0.3406	-0.6884	1.9384	2.6018	3.1851	4.0928	5.7898
		MD2	0.0083	-0.1051	-0.1642	-0.2665	-0.4801	1.3320	1.7881	2.1664	2.7813	3.9387
		MD3	0.2973	0.0694	-0.0347	-0.1637	-0.4923	1.8616	2.5004	3.0480	3.9496	5.6056
		MD4	0.5982	0.3241	0.1942	0.0931	-0.3855	2.0131	2.7150	3.3394	4.2780	5.7357
		MD5	0.7636	0.7256	0.8171	0.9617	1.1642	2.2608	3.0008	3.6723	4.7540	6.7198
	0.02%	MD1	1.7754	0.7654	0.3623	-0.0240	-0.5335	2.6930	2.7658	3.2418	4.1063	5.7931
		MD2	1.0314	0.4247	0.1824	-0.0597	-0.3832	1.7660	1.8755	2.2016	2.7902	3.9396
		MD3	1.7599	0.8255	0.4637	0.1362	-0.3446	2.5947	2.6714	3.1171	3.9706	5.6138
		MD4	2.0993	1.1072	0.7104	0.4038	-0.2292	2.8749	2.9169	3.4103	4.3015	5.7338
		MD5	2.3110	1.5311	1.3481	1.2733	1.3218	3.2132	3.3351	3.8631	4.8561	6.7649
20-sec	0.005%	MD1	-1.5467	-0.9622	-0.7993	-0.7283	-0.8956	2.4372	2.7904	3.2566	4.1020	5.7768
		MD2	-1.1788	-0.7181	-0.5846	-0.5050	-0.5812	1.7536	1.9349	2.2284	2.8114	3.9478
		MD3	-1.3682	-0.8045	-0.6200	-0.5249	-0.6819	2.2428	2.6222	3.0947	3.9504	5.6025
		MD4	-0.9199	-0.4870	-0.3398	-0.2635	-0.5557	2.0994	2.7426	3.3502	4.2524	5.7214
		MD5	-1.0317	-0.2025	0.1975	0.5500	1.0047	2.3152	2.9314	3.5707	4.6778	6.6831
	0.01%	MD1	-0.6734	-0.5496	-0.5349	-0.5673	-0.8201	2.0228	2.6981	3.2233	4.0895	5.7779
		MD2	-0.5992	-0.4502	-0.4083	-0.3976	-0.5312	1.4378	1.8645	2.1985	2.8037	3.9450
		MD3	-0.5402	-0.4180	-0.3659	-0.3737	-0.6084	1.8843	2.5502	3.0716	3.9442	5.6023
		MD4	-0.0651	-0.0843	-0.0740	-0.1068	-0.4749	1.8923	2.7042	3.3363	4.2478	5.7139
		MD5	-0.1532	0.2071	0.4649	0.7054	1.0816	2.1046	2.9532	3.6154	4.7125	6.7001
	0.02%	MD1	1.0071	0.2600	-0.0057	-0.2514	-0.6661	2.2303	2.7010	3.2216	4.0919	5.7790
		MD2	0.5209	0.0848	-0.0555	-0.1874	-0.4308	1.4791	1.8496	2.1860	2.7981	3.9492
		MD3	1.0538	0.3486	0.1407	-0.0751	-0.4653	2.1461	2.5794	3.0836	3.9558	5.6056
		MD4	1.5600	0.7092	0.4478	0.2065	-0.3182	2.4765	2.8021	3.3664	4.2579	5.7108
		MD5	1.5316	1.0172	1.0014	1.0215	1.2331	2.6637	3.1626	3.7584	4.7917	6.7416

Notes: ME and RMSE are of the annualized standard deviation in percentage. The average true daily integrated volatility is around 40% for MD1, 27% for MD2, 36% for MD3, 28% for MD4 and 40% for MD5. MD1 and MD2 are the Heston model at different volatility level. MD3 is the two-factor stochastic volatility model with intraday volatility periodicity. MD4 is the deterministic volatility model with intraday volatility periodicity and MD5 is the Heston model (MD1) with price jumps. The first column indicates the average duration of the observed simulated transactions.

Table B.6: ME and RMSE of daily volatility estimates using the TRV method without subsampling under the BTS scheme

Sparsity	NSR	Model	ME					RMSE				
			1-min	2-min	3-min	5-min	10-min	1-min	2-min	3-min	5-min	10-min
5-sec	0.005%	MD1	-0.7386	-0.0642	0.0913	0.1680	0.1851	1.9218	2.4472	2.9922	3.9446	5.6809
		MD2	-0.6463	-0.1267	-0.0021	0.0488	0.0874	1.3902	1.6765	2.0488	2.6698	3.8747
		MD3	-0.6402	-0.1384	0.0125	0.1026	0.1165	1.7261	2.1962	2.6836	3.4978	5.0122
		MD4	-0.3995	-0.2969	-0.2174	-0.0985	-0.0731	1.4002	1.8229	2.1778	2.7810	3.9711
		MD5	-0.4117	0.3786	0.6509	1.0585	1.7189	1.9117	2.5964	3.2405	4.3322	6.4623
	0.01%	MD1	0.1085	0.3520	0.3439	0.3142	0.2626	1.8267	2.5061	3.0464	3.9573	5.6974
		MD2	-0.0949	0.1486	0.1742	0.1869	0.1258	1.2590	1.7014	2.0785	2.7004	3.8801
		MD3	0.1551	0.2208	0.2532	0.2680	0.1935	1.6491	2.2327	2.6967	3.5108	5.0458
		MD4	0.3948	0.1199	0.0535	0.0674	-0.0045	1.4381	1.8333	2.2001	2.8060	4.0164
		MD5	0.4446	0.8077	0.9476	1.2144	1.7896	1.9562	2.7180	3.3357	4.3813	6.4896
	0.02%	MD1	1.7358	1.1392	0.8667	0.6163	0.3780	2.5885	2.7845	3.2021	4.0384	5.7294
		MD2	0.9993	0.6754	0.5228	0.3796	0.2312	1.6857	1.8834	2.1614	2.7484	3.8905
		MD3	1.7085	1.0048	0.7713	0.5630	0.3422	2.4225	2.4853	2.8486	3.5908	5.0948
		MD4	1.9701	0.9155	0.6078	0.3855	0.1686	2.4677	2.1045	2.3352	2.8921	4.0595
		MD5	2.1045	1.6356	1.4884	1.5421	1.9414	2.9157	3.1288	3.5805	4.5353	6.5625
10-sec	0.005%	MD1	-1.1362	-0.2800	-0.0021	0.1282	0.1350	2.1189	2.4638	2.9893	3.9077	5.6843
		MD2	-0.9108	-0.2736	-0.0638	0.0594	0.0915	1.5407	1.6856	2.0465	2.6697	3.8474
		MD3	-1.0175	-0.3829	-0.1136	0.0598	0.0812	1.9128	2.2274	2.6643	3.4810	5.0171
		MD4	-0.7556	-0.5138	-0.3702	-0.2073	-0.1186	1.5765	1.8882	2.2048	2.8001	3.9377
		MD5	-0.8024	0.1483	0.5708	0.9862	1.7288	2.0396	2.5735	3.2257	4.2856	6.4519
	0.01%	MD1	-0.2648	0.1410	0.2729	0.3009	0.2069	1.8299	2.4731	3.0344	3.9613	5.6767
		MD2	-0.3312	-0.0040	0.1217	0.1688	0.1481	1.2977	1.6873	2.0651	2.6968	3.8538
		MD3	-0.2125	0.0277	0.1589	0.2442	0.1779	1.6604	2.2178	2.7075	3.5036	5.0578
		MD4	0.0357	-0.0954	-0.0744	-0.0169	-0.0381	1.4169	1.8405	2.2106	2.8133	3.9915
		MD5	0.0877	0.6132	0.8651	1.1572	1.7822	1.9288	2.6727	3.3108	4.3771	6.4703
	0.02%	MD1	1.4171	0.9876	0.8335	0.6269	0.3334	2.3837	2.7309	3.1873	4.0413	5.7200
		MD2	0.7940	0.5659	0.4916	0.3681	0.2306	1.5695	1.8326	2.1714	2.7429	3.8912
		MD3	1.3803	0.8448	0.7130	0.5402	0.3251	2.2088	2.4267	2.8389	3.5947	5.1065
		MD4	1.6123	0.7335	0.4714	0.3079	0.1404	2.2035	2.0324	2.2972	2.8741	4.0367
		MD5	1.7771	1.4778	1.4441	1.5273	1.9046	2.6918	3.0593	3.5680	4.5321	6.5540
20-sec	0.005%	MD1	-1.5495	-0.6087	-0.2302	0.0096	-0.0024	2.4113	2.5365	3.0181	3.9089	5.6326
		MD2	-1.1726	-0.4957	-0.2067	-0.0391	-0.0072	1.7383	1.7595	2.0742	2.6700	3.8550
		MD3	-1.4348	-0.6843	-0.3349	-0.0682	0.0042	2.2082	2.3088	2.7062	3.4933	4.9995
		MD4	-1.1779	-0.8037	-0.6032	-0.3917	-0.2569	1.9163	2.0454	2.2940	2.8240	3.9595
		MD5	-1.1827	-0.1329	0.3770	0.8899	1.5604	2.2991	2.6091	3.2144	4.3198	6.4341
	0.01%	MD1	-0.6885	-0.1516	0.0716	0.1681	0.0729	1.9918	2.4967	3.0271	3.9504	5.6821
		MD2	-0.5947	-0.1963	-0.0180	0.0570	0.0278	1.4211	1.7180	2.0724	2.6869	3.8598
		MD3	-0.6264	-0.2355	-0.0294	0.1117	0.0690	1.8182	2.2514	2.7065	3.5107	5.0339
		MD4	-0.4162	-0.3635	-0.3065	-0.2071	-0.1711	1.5767	1.9369	2.2577	2.8166	3.9762
		MD5	-0.3131	0.3306	0.7023	1.0783	1.6439	2.0171	2.6669	3.3037	4.3704	6.5018
	0.02%	MD1	0.9673	0.7576	0.6773	0.4931	0.2356	2.1770	2.6631	3.1650	4.0161	5.7011
		MD2	0.5051	0.4091	0.3836	0.3007	0.1155	1.4548	1.8036	2.1500	2.7316	3.8792
		MD3	0.9340	0.6284	0.5359	0.4205	0.2121	2.0075	2.3824	2.8064	3.5834	5.1012
		MD4	1.1059	0.5034	0.3007	0.1720	0.0337	1.9364	2.0141	2.3149	2.8809	4.0862
		MD5	1.3600	1.2454	1.3184	1.4304	1.7997	2.4864	2.9726	3.5423	4.5144	6.5563

Notes: ME and RMSE are of the annualized standard deviation in percentage. The average true daily integrated volatility is around 40% for MD1, 27% for MD2, 36% for MD3, 28% for MD4 and 40% for MD5. MD1 and MD2 are the Heston model at different volatility level. MD3 is the two-factor stochastic volatility model with intraday volatility periodicity. MD4 is the deterministic volatility model with intraday volatility periodicity and MD5 is the Heston model (MD1) with price jumps. The first column indicates the average duration of the observed simulated transactions.

Table B.7: ME and RMSE of daily volatility estimates using the TRV method with subsampling under the CTS scheme

Sparsity	NSR	Model	ME					RMSE				
			1-min	2-min	3-min	5-min	10-min	1-min	2-min	3-min	5-min	10-min
5-sec	0.005%	MD1	0.2968	0.0332	-0.1110	-0.3157	-0.7328	1.3470	1.8355	2.2431	2.9109	4.1736
		MD2	0.0464	-0.0588	-0.1266	-0.2462	-0.4917	0.9051	1.2467	1.5296	1.9877	2.8346
		MD3	0.2855	0.0506	-0.0949	-0.3215	-0.8718	1.2761	1.7448	2.1428	2.7859	3.9621
		MD4	0.3617	0.1186	-0.0331	-0.3395	-1.1888	1.3787	1.8834	2.3030	2.9214	3.9184
		MD5	0.8233	0.8215	0.8752	0.9928	1.1545	1.7400	2.3406	2.8583	3.6844	5.2112
	0.01%	MD1	1.0698	0.4202	0.1473	-0.1607	-0.6541	1.6907	1.8630	2.2249	2.8801	4.1466
		MD2	0.5617	0.1989	0.0455	-0.1427	-0.4400	1.0791	1.2513	1.5122	1.9659	2.8164
		MD3	1.0234	0.4194	0.1516	-0.1723	-0.7967	1.5991	1.7737	2.1267	2.7559	3.9334
		MD4	1.1287	0.5029	0.2245	-0.1830	-1.1079	1.7312	1.9213	2.2907	2.8879	3.8802
		MD5	1.6017	1.2112	1.1359	1.1498	1.2340	2.2173	2.4900	2.9326	3.7154	5.2177
	0.02%	MD1	2.5864	1.1887	0.6611	0.1485	-0.4973	2.9151	2.1454	2.2829	2.8455	4.0972
		MD2	1.5734	0.7087	0.3868	0.0631	-0.3363	1.8739	1.4190	1.5412	1.9392	2.7833
		MD3	2.4694	1.1504	0.6410	0.1240	-0.6466	2.7515	2.0426	2.1817	2.7200	3.8799
		MD4	2.6275	1.2649	0.7371	0.1286	-0.9467	2.9351	2.2114	2.3560	2.8479	3.8090
		MD5	3.1305	1.9841	1.6540	1.4628	1.3917	3.5031	2.9285	3.1449	3.7972	5.2348
10-sec	0.005%	MD1	-0.1986	-0.0693	-0.1605	-0.3369	-0.7443	1.3738	1.8617	2.2630	2.9232	4.1799
		MD2	-0.2806	-0.1265	-0.1592	-0.2615	-0.5001	0.9729	1.2699	1.5451	1.9959	2.8382
		MD3	-0.1666	-0.0480	-0.1414	-0.3473	-0.8851	1.3000	1.7749	2.1625	2.7984	3.9685
		MD4	0.0062	0.0435	-0.0719	-0.3596	-1.1973	1.3743	1.9063	2.3208	2.9327	3.9241
		MD5	0.3180	0.7174	0.8221	0.9677	1.1403	1.5978	2.3272	2.8530	3.6881	5.2092
	0.01%	MD1	0.5931	0.3192	0.0976	-0.1820	-0.6660	1.4841	1.8743	2.2426	2.8932	4.1538
		MD2	0.2478	0.1332	0.0129	-0.1578	-0.4484	0.9771	1.2632	1.5277	1.9745	2.8199
		MD3	0.5889	0.3225	0.1055	-0.1991	-0.8103	1.4141	1.7890	2.1445	2.7689	3.9401
		MD4	0.7929	0.4314	0.1863	-0.2033	-1.1160	1.5752	1.9329	2.3062	2.8995	3.8864
		MD5	1.1170	1.1088	1.0847	1.1244	1.2191	1.9286	2.4652	2.9259	3.7205	5.2165
	0.02%	MD1	2.1409	1.0900	0.6120	0.1277	-0.5098	2.5611	2.1284	2.2948	2.8594	4.1057
		MD2	1.2809	0.6472	0.3555	0.0481	-0.3451	1.6459	1.4147	1.5539	1.9481	2.7873
		MD3	2.0635	1.0573	0.5956	0.0964	-0.6603	2.4335	2.0308	2.1953	2.7352	3.8878
		MD4	2.3179	1.1985	0.6992	0.1084	-0.9546	2.6907	2.2043	2.3668	2.8602	3.8159
		MD5	2.6775	1.8878	1.6048	1.4370	1.3775	3.1293	2.8890	3.1360	3.8034	5.2342
20-sec	0.005%	MD1	-3.3553	-0.5888	-0.3606	-0.4217	-0.7800	3.7019	2.0158	2.3321	2.9616	4.2012
		MD2	-2.3673	-0.4691	-0.2902	-0.3118	-0.5229	2.6247	1.3902	1.5939	2.0268	2.8530
		MD3	-3.0215	-0.5208	-0.3260	-0.4202	-0.9125	3.3293	1.8981	2.2234	2.8353	3.9885
		MD4	-2.2783	-0.3384	-0.2228	-0.4241	-1.2220	2.7230	1.9950	2.3731	2.9674	3.9414
		MD5	-2.8934	0.1894	0.6246	0.8867	1.1063	3.3754	2.2727	2.8414	3.6921	5.2141
	0.01%	MD1	-2.5748	-0.1880	-0.0983	-0.2653	-0.7011	2.9972	1.9304	2.2952	2.9308	4.1765
		MD2	-1.8486	-0.2026	-0.1163	-0.2079	-0.4711	2.1414	1.3196	1.5647	2.0046	2.8359
		MD3	-2.2816	-0.1378	-0.0753	-0.2710	-0.8380	2.6767	1.8243	2.1903	2.8057	3.9609
		MD4	-1.5119	0.0588	0.0394	-0.2671	-1.1411	2.1076	1.9520	2.3457	2.9330	3.9030
		MD5	-2.1120	0.5936	0.8898	1.0450	1.1849	2.7219	2.3378	2.9036	3.7237	5.2218
	0.02%	MD1	-1.0688	0.6032	0.4232	0.0452	-0.5451	1.8517	2.0081	2.3156	2.8965	4.1306
		MD2	-0.8447	0.3238	0.2298	-0.0002	-0.3675	1.3346	1.3489	1.5690	1.9776	2.8039
		MD3	-0.8505	0.6145	0.4210	0.0258	-0.6879	1.6498	1.9103	2.2098	2.7699	3.9106
		MD4	-0.0426	0.8431	0.5573	0.0443	-0.9808	1.4504	2.1023	2.3809	2.8916	3.8335
		MD5	-0.5983	1.3892	1.4158	1.3596	1.3408	1.8062	2.6532	3.0935	3.8061	5.2409

Notes: ME and RMSE are of the annualized standard deviation in percentage. The average true daily integrated volatility is around 40% for MD1, 27% for MD2, 36% for MD3, 28% for MD4 and 40% for MD5. MD1 and MD2 are the Heston model at different volatility level. MD3 is the two-factor stochastic volatility model with intraday volatility periodicity. MD4 is the deterministic volatility model with intraday volatility periodicity and MD5 is the Heston model (MD1) with price jumps. The first column indicates the average duration of the observed simulated transactions.

Table B.8: ME and RMSE of daily volatility estimates using the TRV method with subsampling under the TTS scheme

Sparsity	NSR	Model	ME					RMSE				
			1-min	2-min	3-min	5-min	10-min	1-min	2-min	3-min	5-min	10-min
5-sec	0.005%	MD1	-0.0517	-0.1758	-0.2597	-0.4047	-0.7809	1.3130	1.8416	2.2588	2.9293	4.1866
		MD2	-0.1824	-0.1986	-0.2244	-0.3025	-0.5233	0.9132	1.2584	1.5419	1.9989	2.8422
		MD3	-0.0293	-0.1394	-0.2269	-0.4029	-0.9162	1.2428	1.7475	2.1521	2.7986	3.9739
		MD4	0.1113	-0.0348	-0.1454	-0.4131	-1.2243	1.3307	1.8849	2.3105	2.9296	3.9347
		MD5	0.4782	0.6174	0.7243	0.8982	1.1101	1.6043	2.2772	2.8182	3.6622	5.1995
	0.01%	MD1	0.7340	0.2171	0.0021	-0.2481	-0.7017	1.4892	1.8218	2.2209	2.8926	4.1580
		MD2	0.3415	0.0625	-0.0507	-0.1984	-0.4712	0.9629	1.2311	1.5114	1.9728	2.8230
		MD3	0.7199	0.2349	0.0223	-0.2529	-0.8404	1.4171	1.7349	2.1186	2.7628	3.9436
		MD4	0.8908	0.3555	0.1154	-0.2555	-1.1430	1.5773	1.8896	2.2835	2.8914	3.8950
		MD5	1.2689	1.0131	0.9879	1.0568	1.1894	1.9802	2.3968	2.8792	3.6887	5.2048
	0.02%	MD1	2.2717	0.9948	0.5212	0.0635	-0.5444	2.6235	2.0319	2.2421	2.8465	4.1057
		MD2	1.3660	0.5791	0.2948	0.0090	-0.3670	1.6798	1.3472	1.5159	1.9382	2.7879
		MD3	2.1836	0.9759	0.5173	0.0456	-0.6897	2.4897	1.9411	2.1421	2.7169	3.8877
		MD4	2.4054	1.1268	0.6331	0.0588	-0.9813	2.7283	2.1318	2.3233	2.8417	3.8208
		MD5	2.8163	1.7964	1.5130	1.3722	1.3472	3.2105	2.7970	3.0698	3.7630	5.2200
10-sec	0.005%	MD1	-0.5516	-0.4487	-0.4361	-0.5224	-0.8358	1.4790	1.9039	2.2990	2.9580	4.2085
		MD2	-0.5129	-0.3793	-0.3494	-0.3808	-0.5574	1.0656	1.3125	1.5773	2.0197	2.8519
		MD3	-0.4819	-0.3884	-0.4015	-0.5122	-0.9712	1.3894	1.8081	2.1963	2.8303	3.9928
		MD4	-0.2505	-0.2400	-0.2896	-0.5027	-1.2683	1.4059	1.9158	2.3326	2.9460	3.9521
		MD5	-0.0248	0.3341	0.5398	0.7829	1.0467	1.5891	2.2287	2.7879	3.6465	5.1956
	0.01%	MD1	0.2364	-0.0451	-0.1707	-0.3643	-0.7566	1.3765	1.8287	2.2422	2.9157	4.1797
		MD2	0.0132	-0.1109	-0.1730	-0.2760	-0.5052	0.9300	1.2483	1.5338	1.9901	2.8323
		MD3	0.2681	-0.0040	-0.1493	-0.3617	-0.8952	1.3166	1.7460	2.1455	2.7892	3.9618
		MD4	0.5300	0.1618	-0.0235	-0.3432	-1.1865	1.4605	1.8814	2.2906	2.9031	3.9119
		MD5	0.7703	0.7409	0.8070	0.9428	1.1259	1.7564	2.3086	2.8352	3.6692	5.2007
	0.02%	MD1	1.7736	0.7541	0.3566	-0.0494	-0.5989	2.2414	1.9473	2.2272	2.8581	4.1262
		MD2	1.0392	0.4206	0.1775	-0.0671	-0.4010	1.4356	1.3033	1.5124	1.9480	2.7963
		MD3	1.7329	0.7567	0.3539	-0.0612	-0.7437	2.1514	1.8708	2.1337	2.7332	3.9052
		MD4	2.0392	0.9541	0.5017	-0.0268	-1.0242	2.4452	2.0641	2.3026	2.8455	3.8372
		MD5	2.3202	1.5468	1.3390	1.2604	1.2850	2.8184	2.6578	3.0024	3.7363	5.2146
20-sec	0.005%	MD1	-1.5456	-0.9755	-0.8101	-0.7508	-0.9624	2.1294	2.1667	2.4446	3.0322	4.2458
		MD2	-1.1754	-0.7228	-0.5899	-0.5349	-0.6353	1.5498	1.5021	1.6784	2.0747	2.8870
		MD3	-1.3862	-0.8709	-0.7288	-0.7174	-1.0831	1.9469	2.0256	2.3181	2.8971	4.0378
		MD4	-0.9656	-0.6310	-0.5652	-0.6721	-1.3640	1.7472	2.0611	2.4288	3.0082	3.9960
		MD5	-1.0312	-0.1876	0.1763	0.5540	0.9132	1.9665	2.2887	2.7769	3.6325	5.1833
	0.01%	MD1	-0.6723	-0.5651	-0.5403	-0.5906	-0.8826	1.5976	1.9994	2.3519	2.9801	4.2153
		MD2	-0.5942	-0.4507	-0.4112	-0.4286	-0.5833	1.1564	1.3784	1.6107	2.0385	2.8665
		MD3	-0.5574	-0.4809	-0.4722	-0.5639	-1.0070	1.4725	1.8769	2.2341	2.8474	4.0056
		MD4	-0.1095	-0.2245	-0.2953	-0.5106	-1.2820	1.4410	1.9525	2.3592	2.9575	3.9539
		MD5	-0.1518	0.2245	0.4473	0.7158	0.9933	1.6700	2.2787	2.7934	3.6475	5.1874
	0.02%	MD1	1.0095	0.2420	-0.0044	-0.2698	-0.7231	1.7813	1.9083	2.2582	2.9023	4.1593
		MD2	0.5272	0.0861	-0.0554	-0.2161	-0.4775	1.1592	1.2928	1.5381	1.9830	2.8277
		MD3	1.0376	0.2876	0.0374	-0.2591	-0.8542	1.7195	1.8116	2.1543	2.7751	3.9461
		MD4	1.5196	0.5736	0.2358	-0.1903	-1.1188	2.0884	1.9865	2.3133	2.8833	3.8762
		MD5	1.5376	1.0383	0.9871	1.0383	1.1526	2.2818	2.4775	2.9055	3.7002	5.1988

Notes: ME and RMSE are of the annualized standard deviation in percentage. The average true daily integrated volatility is around 40% for MD1, 27% for MD2, 36% for MD3, 28% for MD4 and 40% for MD5. MD1 and MD2 are the Heston model at different volatility level. MD3 is the two-factor stochastic volatility model with intraday volatility periodicity. MD4 is the deterministic volatility model with intraday volatility periodicity and MD5 is the Heston model (MD1) with price jumps. The first column indicates the average duration of the observed simulated transactions.

Table B.9: ME and RMSE of daily volatility using the TRV method with subsampling under the BTS scheme

Sparsity	NSR	Model	ME					RMSE				
			1-min	2-min	3-min	5-min	10-min	1-min	2-min	3-min	5-min	10-min
5-sec	0.005%	MD1	-0.4286	0.0470	0.0031	0.0004	-0.1777	1.3869	1.7228	2.1610	2.8165	4.0975
		MD2	-0.4321	-0.0500	-0.0513	-0.0338	-0.1309	0.9964	1.1706	1.4665	1.9242	2.7797
		MD3	-0.4435	0.0200	0.0123	-0.0019	-0.1916	1.2425	1.5657	1.9431	2.5511	3.6823
		MD4	-0.2403	-0.0913	0.0126	0.0055	-0.2069	1.0574	1.3723	1.6849	2.2278	3.2164
		MD5	-0.0687	0.6282	0.7732	1.0940	1.6463	1.4163	2.0196	2.5587	3.4485	5.2161
	0.01%	MD1	0.3382	0.4486	0.2625	0.1532	-0.1142	1.3520	1.7646	2.1562	2.8077	4.0850
		MD2	0.0786	0.2172	0.1192	0.0675	-0.0884	0.9049	1.1831	1.4608	1.9152	2.7695
		MD3	0.3291	0.4002	0.2555	0.1439	-0.1301	1.2126	1.6016	1.9521	2.5460	3.6734
		MD4	0.5904	0.3128	0.2734	0.1532	-0.1474	1.1809	1.4043	1.7059	2.2373	3.2230
		MD5	0.7280	1.0331	1.0365	1.2492	1.7128	1.5989	2.1671	2.6391	3.4936	5.2306
	0.02%	MD1	1.8463	1.2523	0.7789	0.4499	0.0126	2.2762	2.1053	2.2506	2.8148	4.0590
		MD2	1.0802	0.7479	0.4603	0.2651	-0.0059	1.4503	1.3864	1.5109	1.9145	2.7529
		MD3	1.7862	1.1644	0.7397	0.4265	-0.0167	2.1440	1.9166	2.0542	2.5558	3.6603
		MD4	2.2616	1.1036	0.7889	0.4437	-0.0291	2.4966	1.7643	1.8609	2.2798	3.2357
		MD5	2.2713	1.8503	1.5649	1.5559	1.8424	2.7080	2.6518	2.8753	3.5982	5.2587
10-sec	0.005%	MD1	-0.9375	-0.0435	-0.0307	-0.0333	-0.1880	1.6335	1.8357	2.1600	2.8457	4.1021
		MD2	-0.8123	-0.1034	-0.0778	-0.0560	-0.1417	1.2304	1.2500	1.4735	1.9394	2.7833
		MD3	-0.8436	-0.0958	-0.0423	-0.0340	-0.2088	1.4854	1.6584	1.9742	2.5751	3.7013
		MD4	-0.5797	-0.2436	-0.0830	-0.0584	-0.2432	1.2715	1.4404	1.7121	2.2547	3.2308
		MD5	-0.5986	0.5429	0.7290	1.0637	1.6334	1.5659	2.0991	2.5573	3.4711	5.2171
	0.01%	MD1	0.0713	0.3397	0.2356	0.1224	-0.1251	1.3501	1.8561	2.1611	2.8355	4.0856
		MD2	-0.1345	0.1520	0.1010	0.0460	-0.0992	0.9467	1.2544	1.4661	1.9323	2.7757
		MD3	0.1028	0.2815	0.2097	0.1090	-0.1529	1.2390	1.6800	1.9774	2.5686	3.6909
		MD4	0.3144	0.1905	0.1929	0.0949	-0.1830	1.1801	1.4321	1.7293	2.2613	3.2356
		MD5	0.4190	0.9362	0.9980	1.2244	1.6968	1.5275	2.2311	2.6363	3.5126	5.2337
	0.02%	MD1	2.0479	1.1137	0.7770	0.4235	-0.0051	2.5005	2.1285	2.2661	2.8358	4.0650
		MD2	1.1815	0.6675	0.4586	0.2426	-0.0175	1.5885	1.4198	1.5277	1.9301	2.7608
		MD3	1.9622	1.0222	0.7158	0.3933	-0.0340	2.3465	1.9448	2.0766	2.5855	3.6804
		MD4	2.1082	1.0387	0.7207	0.3933	-0.0644	2.4205	1.7728	1.8823	2.2997	3.2520
		MD5	2.4111	1.7259	1.5437	1.5352	1.8277	2.8718	2.6669	2.8772	3.6225	5.2623
20-sec	0.005%	MD1	-1.9043	-0.4650	-0.1715	-0.1586	-0.3052	2.5154	1.9081	2.3632	2.8884	4.1176
		MD2	-1.4513	-0.3972	-0.1718	-0.1551	-0.2189	1.8465	1.3232	1.6142	1.9728	2.8010
		MD3	-1.7349	-0.4945	-0.1976	-0.1617	-0.3245	2.2880	1.7421	2.0996	2.6517	3.7326
		MD4	-1.3673	-0.5413	-0.2730	-0.2261	-0.3586	1.9372	1.6463	1.8101	2.3044	3.2759
		MD5	-1.5684	0.0539	0.6149	0.9467	1.5263	2.3452	2.0185	2.7000	3.4925	5.2178
	0.01%	MD1	-0.7727	0.0595	0.0795	-0.0012	-0.2474	1.8388	1.8658	2.3619	2.8691	4.1079
		MD2	-0.6930	-0.0470	-0.0025	-0.0515	-0.1798	1.3436	1.2747	1.6111	1.9649	2.7898
		MD3	-0.6715	0.0110	0.0681	-0.0090	-0.2678	1.6684	1.6857	2.1131	2.6393	3.7245
		MD4	-0.3601	-0.0577	0.0292	-0.0623	-0.3056	1.4354	1.5473	1.7854	2.3140	3.2870
		MD5	-0.4358	0.5872	0.8854	1.1072	1.5868	1.8245	2.1169	2.7876	3.5323	5.2285
	0.02%	MD1	1.4062	1.0834	0.5790	0.3089	-0.1229	2.2543	2.1942	2.4346	2.8722	4.0765
		MD2	0.7467	0.6389	0.3302	0.1612	-0.0953	1.4646	1.4674	1.6470	1.9576	2.7740
		MD3	1.3598	0.9923	0.5629	0.2802	-0.1573	2.1106	1.9812	2.2102	2.6445	3.7064
		MD4	1.6100	0.9172	0.6508	0.2535	-0.1811	2.1813	1.8069	1.9153	2.3599	3.2984
		MD5	1.7450	1.6299	1.4063	1.4237	1.7176	2.5623	2.6474	3.0191	3.6372	5.2592

Notes: ME and RMSE are of the annualized standard deviation in percentage. The average true daily integrated volatility is around 40% for MD1, 27% for MD2, 36% for MD3, 28% for MD4 and 40% for MD5. MD1 and MD2 are the Heston model at different volatility level. MD3 is the two-factor stochastic volatility model with intraday volatility periodicity. MD4 is the deterministic volatility model with intraday volatility periodicity and MD5 is the Heston model (MD1) with price jumps. The first column indicates the average duration of the observed simulated transactions. The sampling frequency of the BTS scheme equals to twice the average transaction duration.

Table B.10: ME and RMSE of daily volatility estimates of the RK and ACD-ICV methods for Model MD2

Sparsity	NSR	ME							RMSE						
		RK	ACD-ICV	Avg. sampling frequency (ACD-ICV)					RK	ACD-ICV	Avg. sampling frequency (ACD-ICV)				
				1-min	3-min	5-min	10-min	15-min			1-min	3-min	5-min	10-min	15-min
5-sec	0.005%	-0.0554	ME1	-3.7334	-2.1773	-1.6815	-1.1075	-0.8023	1.5658	ME1	4.2642	2.6916	2.3084	2.1126	2.1477
			ME2	-0.0282	-0.0874	-0.0893	-0.0093	0.0975		ME2	1.6114	1.4167	1.5023	1.7761	1.9821
			ME3	-0.1101	-0.0087	0.1370	0.2380	0.4272		ME3	0.8800	0.9494	1.1542	0.9745	1.0777
	0.01%	-0.0417	ME1	-2.1706	-1.2209	-0.9427	-0.6030	-0.3953	1.5672	ME1	2.7892	1.8687	1.7655	1.8571	2.0231
			ME2	0.0931	0.0477	0.0556	0.1425	0.2551		ME2	1.5181	1.3601	1.4705	1.7590	1.9984
			ME3	0.0629	0.1609	0.3026	0.4080	0.5997		ME3	0.8433	0.9261	1.1347	1.0027	1.1378
	0.02%	-0.0147	ME1	1.1040	0.5697	0.3923	0.3011	0.3363	1.5703	ME1	1.8511	1.4601	1.5144	1.7720	2.0300
			ME2	0.3425	0.3309	0.3546	0.4534	0.5803		ME2	1.4140	1.3708	1.5045	1.8071	2.0864
			ME3	0.4075	0.4979	0.6377	0.7500	0.9453		ME3	0.8902	0.9884	1.2067	1.1571	1.3336
10-sec	0.005%	-0.0751	ME1	-7.1362	-3.9157	-2.9626	-1.9998	-1.5156	1.7937	ME1	7.7312	4.3933	3.4815	2.7513	2.5519
			ME2	-0.0467	-0.1531	-0.1661	-0.1066	0.0051		ME2	2.0058	1.5586	1.5798	1.7972	2.0045
			ME3	-0.1574	-0.1860	-0.0314	0.1258	0.3150		ME3	1.0752	0.9808	1.0817	1.0054	1.0842
	0.01%	-0.0599	ME1	-5.7284	-3.1380	-2.3711	-1.5900	-1.1781	1.7954	ME1	6.2967	3.6246	2.9215	2.4286	2.3437
			ME2	0.0923	0.0009	-0.0105	0.0578	0.1712		ME2	1.8501	1.4889	1.5312	1.7770	2.0051
			ME3	0.0175	-0.0083	0.1497	0.2992	0.4906		ME3	1.0274	0.9357	1.0661	1.0207	1.1304
	0.02%	-0.0294	ME1	-3.1350	-1.6696	-1.2830	-0.8363	-0.5822	1.7983	ME1	3.7327	2.2514	2.0041	1.9572	2.0764
			ME2	0.4040	0.3022	0.3028	0.3849	0.4983		ME2	1.6709	1.4263	1.5193	1.8024	2.0552
			ME3	0.3654	0.3408	0.4923	0.6466	0.8419		ME3	1.0312	0.9639	1.1165	1.1487	1.3071
20-sec	0.005%	-0.1078	ME1	-11.3529	-6.1908	-4.7306	-3.1880	-2.4821	2.0838	ME1	12.1410	6.7322	5.2491	3.8280	3.3078
			ME2	-0.1409	-0.3073	-0.3393	-0.2906	-0.1960		ME2	2.6834	1.8420	1.7730	1.9061	2.0670
			ME3	-0.2602	-0.3610	-0.2607	-0.1048	0.0867		ME3	1.7642	1.1645	1.2405	1.0845	1.1195
	0.01%	-0.0907	ME1	-10.1662	-5.6344	-4.2923	-2.8841	-2.2187	2.0856	ME1	10.9460	6.1576	4.7993	3.5372	3.0841
			ME2	0.0236	-0.1482	-0.1796	-0.1280	-0.0246		ME2	2.5716	1.7486	1.6904	1.8563	2.0394
			ME3	-0.0823	-0.1821	-0.0793	0.0740	0.2668		ME3	1.7119	1.0965	1.1938	1.0646	1.1348
	0.02%	-0.0571	ME1	-8.3626	-4.5732	-3.4209	-2.2876	-1.7304	2.0893	ME1	9.0708	5.0768	3.9382	3.0069	2.7036
			ME2	0.3623	0.1767	0.1515	0.2106	0.3181		ME2	2.4080	1.6406	1.6133	1.8301	2.0390
			ME3	0.2731	0.1757	0.2878	0.4300	0.6247		ME3	1.6765	1.0623	1.2123	1.1294	1.2590

Notes: ME and RMSE are of the annualized standard deviation in percentage. MD2 is the Heston model and the average true daily integrated volatility is around 27%. ME1 is the ACD-ICV method in Tse and Yang (2012). ME2 is ME1 with δ^2 replaced by V_D, the integrated volatility estimated using TRV with subsampling at 3-min sampling frequency. ME3 is ME2 with sampled durations computed from BTS returns. All ACD models are fitted to diurnally transformed durations using the time-transformation function based on the number of trades as in Tse and Dong (2014).

Table B.11: ME and RMSE of daily volatility estimates of the RK and ACD-ICV methods for Model MD3

Sparsity	NSR	ME							RMSE						
		RK	ACD-ICV	Avg. sampling frequency (ACD-ICV)					RK	ACD-ICV	Avg. sampling frequency (ACD-ICV)				
				1-min	3-min	5-min	10-min	15-min			1-min	3-min	5-min	10-min	15-min
5-sec	0.005%	-0.0985	ME1	-5.1430	-2.9821	-2.2665	-1.6531	-1.5058	2.1510	ME1	5.2778	3.1797	2.5494	2.3407	2.6393
			ME2	-0.1392	-0.0764	-0.0110	-0.0779	-0.2105		ME2	1.1546	1.0977	1.1627	1.7015	2.2357
			ME3	-0.1278	-0.0437	0.0464	0.2487	0.4009		ME3	1.1012	1.1052	1.0151	1.0186	1.1564
	0.01%	-0.0784	ME1	-2.8997	-1.6472	-1.2278	-0.9581	-0.9320	2.1529	ME1	3.1170	1.9593	1.6827	1.9562	2.3715
			ME2	0.1041	0.1734	0.2389	0.1467	0.0340		ME2	1.1343	1.0815	1.1765	1.7429	2.2203
			ME3	0.1213	0.2033	0.2958	0.4998	0.6496		ME3	1.0659	1.1191	1.0287	1.0839	1.2589
	0.02%	-0.0389	ME1	1.7384	0.8618	0.6599	0.2937	0.0912	2.1569	ME1	2.0370	1.3384	1.3060	1.8157	2.2471
			ME2	0.6048	0.6700	0.7349	0.6094	0.5208		ME2	1.2157	1.2370	1.3525	1.9015	2.3172
			ME3	0.6152	0.6992	0.7930	1.0000	1.1425		ME3	1.1817	1.2374	1.2195	1.3586	1.5773
10-sec	0.005%	-0.1400	ME1	-9.8054	-5.3835	-4.0509	-2.9244	-2.4936	2.4629	ME1	9.9038	5.5136	4.2308	3.3837	3.2661
			ME2	-0.3045	-0.2414	-0.1779	-0.3024	-0.3838		ME2	1.2664	1.1600	1.2053	1.8027	2.2380
			ME3	-0.2861	-0.2256	-0.1295	0.0859	0.2311		ME3	1.0727	1.3510	1.2577	1.0275	1.1793
	0.01%	-0.1176	ME1	-7.5214	-4.2769	-3.2264	-2.3623	-2.0371	2.4649	ME1	7.6449	4.4290	3.4421	2.9256	2.9415
			ME2	-0.0509	0.0108	0.0731	-0.0564	-0.1365		ME2	1.2023	1.1152	1.1856	1.7991	2.2139
			ME3	-0.0344	0.0267	0.1246	0.3414	0.4829		ME3	1.0101	1.2965	1.2193	1.0621	1.2583
	0.02%	-0.0736	ME1	-4.0344	-2.2013	-1.7003	-1.3148	-1.1934	2.4688	ME1	4.1991	2.4601	2.0715	2.2143	2.4857
			ME2	0.4531	0.5178	0.5706	0.4320	0.3584		ME2	1.2043	1.2071	1.3152	1.8937	2.2829
			ME3	0.4672	0.5320	0.6299	0.8471	0.9877		ME3	1.0770	1.3437	1.3165	1.2903	1.5195
20-sec	0.005%	-0.2100	ME1	-15.9076	-8.5820	-6.5409	-4.6350	-3.8335	2.8522	ME1	15.9986	8.6873	6.6717	4.9348	4.3461
			ME2	-0.6112	-0.5503	-0.4999	-0.6579	-0.6963		ME2	1.4445	1.3178	1.3434	1.9301	2.2985
			ME3	-0.5833	-0.5374	-0.4419	-0.2247	-0.0804		ME3	1.1720	1.4465	1.3753	1.0861	1.2128
	0.01%	-0.1848	ME1	-14.3233	-7.7133	-5.8979	-4.1974	-3.4686	2.8539	ME1	14.4451	7.8232	6.0400	4.5338	4.0390
			ME2	-0.3499	-0.2930	-0.2487	-0.4073	-0.4497		ME2	1.3487	1.2193	1.2737	1.8768	2.2533
			ME3	-0.3279	-0.2795	-0.1833	0.0347	0.1712		ME3	1.0570	1.3513	1.2855	1.0519	1.2480
	0.02%	-0.1354	ME1	-11.3285	-6.1740	-4.6783	-3.3754	-2.8060	2.8581	ME1	11.4276	6.3005	4.8518	3.7987	3.5021
			ME2	0.1581	0.2138	0.2440	0.0920	0.0634		ME2	1.2626	1.1887	1.2912	1.8650	2.2347
			ME3	0.1806	0.2296	0.3295	0.5495	0.6822		ME3	1.0035	1.3122	1.2790	1.1717	1.4295

Notes: ME and RMSE are of the annualized standard deviation in percentage. MD3 is the two-factor stochastic volatility model with intraday volatility periodicity and the average true daily integrated volatility is around 36%. ME1 is the ACD-ICV method in Tse and Yang (2012). ME2 is ME1 with δ^2 replaced by V_D, the integrated volatility estimated using TRV with subsampling at 3-min sampling frequency. ME3 is ME2 with sampled durations computed from BTS returns. All ACD models are fitted to diurnally transformed durations using the time-transformation function based on the number of trades as in Tse and Dong (2014).

Table B.12: ME and RMSE of daily volatility estimates of the RK and ACD-ICV methods for Model MD4

Sparsity	NSR	ME							RMSE						
		RK	ACD-ICV	Avg. sampling frequency (ACD-ICV)					RK	ACD-ICV	Avg. sampling frequency (ACD-ICV)				
				1-min	3-min	5-min	10-min	15-min			1-min	3-min	5-min	10-min	15-min
5-sec	0.005%	-0.1392	ME1	-4.3273	-2.7901	-2.1761	-1.4735	-1.0744	2.2287	ME1	4.6580	3.1543	2.6755	2.4237	2.5051
			ME2	-0.0495	-0.0064	0.0371	0.1388	0.2862		ME2	1.5681	1.3739	1.4959	1.9008	2.2658
			ME3	0.0241	0.0089	0.0718	0.2014	0.2764		ME3	1.0547	1.3115	1.3079	1.2890	1.3706
	0.01%	-0.1186	ME1	-2.1935	-1.4121	-1.1154	-0.7545	-0.4749	2.2299	ME1	2.6423	1.9116	1.8335	2.0100	2.2642
			ME2	0.2214	0.2643	0.2917	0.4040	0.5553		ME2	1.4576	1.2903	1.4621	1.8925	2.2714
			ME3	0.2804	0.2694	0.3339	0.4596	0.5391		ME3	1.0469	1.2900	1.3068	1.3242	1.4129
	0.02%	-0.0784	ME1	2.6623	1.2371	0.8316	0.5434	0.5493	2.2332	ME1	3.0175	1.6887	1.5544	1.8622	2.2316
			ME2	0.7270	0.7673	0.8013	0.9268	1.0855		ME2	1.4677	1.3979	1.5569	2.0100	2.4160
			ME3	0.7908	0.7864	0.8547	0.9708	1.0650		ME3	1.2255	1.4206	1.4601	1.5398	1.6284
10-sec	0.005%	-0.1969	ME1	-8.3784	-4.9696	-3.8431	-2.6425	-2.0280	2.5340	ME1	8.6674	5.2711	4.2264	3.3330	3.0782
			ME2	-0.1721	-0.1334	-0.1027	0.0030	0.1455		ME2	1.9351	1.5489	1.6187	1.9670	2.2838
			ME3	-0.0688	-0.0957	-0.0495	0.0660	0.1438		ME3	1.1664	1.1820	1.2376	1.3347	1.4051
	0.01%	-0.1737	ME1	-6.0018	-3.8539	-2.9829	-2.0536	-1.5586	2.5354	ME1	6.3235	4.1662	3.4131	2.8460	2.7661
			ME2	0.1070	0.1400	0.1735	0.2746	0.4205		ME2	1.6517	1.4234	1.5607	1.9354	2.2908
			ME3	0.1925	0.1621	0.2169	0.3401	0.4135		ME3	1.1288	1.1714	1.2203	1.3236	1.4153
	0.02%	-0.1282	ME1	-2.6725	-1.6348	-1.3508	-0.9712	-0.6781	2.5382	ME1	3.1174	2.1176	2.0149	2.1226	2.3370
			ME2	0.6470	0.6619	0.6871	0.8054	0.9596		ME2	1.6232	1.4529	1.6017	2.0189	2.4090
			ME3	0.7104	0.6773	0.7394	0.8777	0.9437		ME3	1.2502	1.3162	1.3717	1.4739	1.6046
20-sec	0.005%	-0.2998	ME1	-13.3204	-7.8200	-6.1131	-4.2413	-3.3202	2.9102	ME1	13.6066	8.1128	6.4488	4.7699	4.1023
			ME2	-0.4123	-0.3825	-0.3583	-0.2556	-0.1106		ME2	2.3759	1.8500	1.8392	2.0798	2.3450
			ME3	-0.2791	-0.3033	-0.2788	-0.1795	-0.1061		ME3	1.5569	1.2135	1.2524	1.3578	1.4859
	0.01%	-0.2743	ME1	-11.3275	-6.8981	-5.4315	-3.7721	-2.9385	2.9106	ME1	11.6507	7.1867	5.7795	4.3379	3.7789
			ME2	-0.1303	-0.1052	-0.0809	0.0211	0.1660		ME2	2.1932	1.6925	1.7505	2.0315	2.3168
			ME3	-0.0116	-0.0343	-0.0073	0.0982	0.1684		ME3	1.4968	1.1375	1.1847	1.3111	1.4504
	0.02%	-0.2249	ME1	-8.9181	-5.1741	-4.0973	-2.8888	-2.2279	2.9122	ME1	9.2785	5.4984	4.4887	3.5481	3.2275
			ME2	0.3752	0.4125	0.4484	0.5575	0.7060		ME2	2.2374	1.6579	1.7088	2.0393	2.3828
			ME3	0.5157	0.4939	0.5281	0.6425	0.7192		ME3	1.5317	1.1882	1.2484	1.3981	1.5441

Notes: ME and RMSE are of the annualized standard deviation in percentage. MD4 is the deterministic volatility model with intraday volatility periodicity and the average true daily integrated volatility is around 28%. ME1 is the ACD-ICV method in Tse and Yang (2012). ME2 is ME1 with δ^2 replaced by V_D, the integrated volatility estimated using TRV with subsampling at 3-min sampling frequency. ME3 is ME2 with sampled durations computed from BTS returns. All ACD models are fitted to diurnally transformed durations using the time-transformation function based on the number of trades as in Tse and Dong (2014).

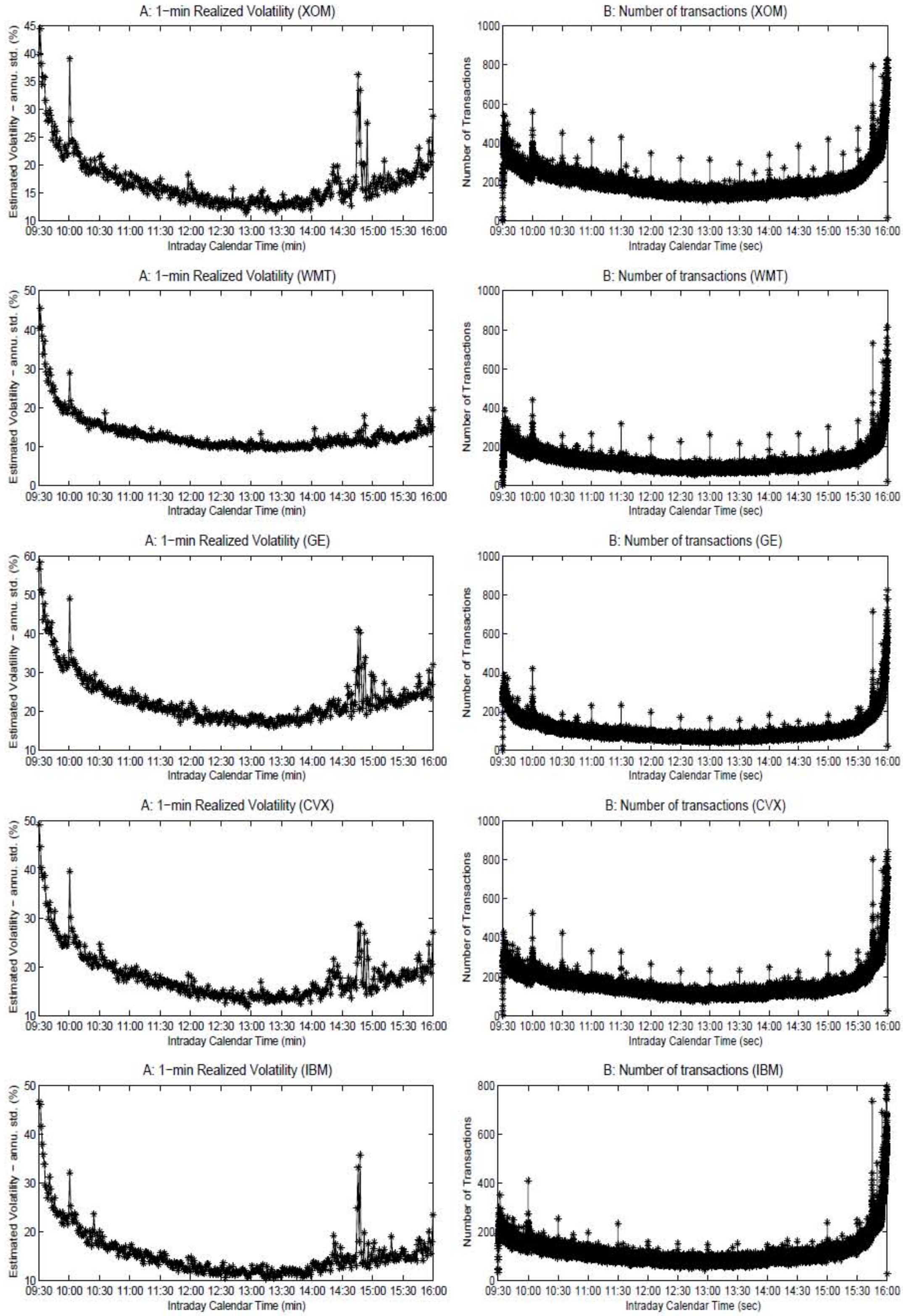

Figure B.1: 1-min Realized Volatility and number of transactions, 2010-2013.

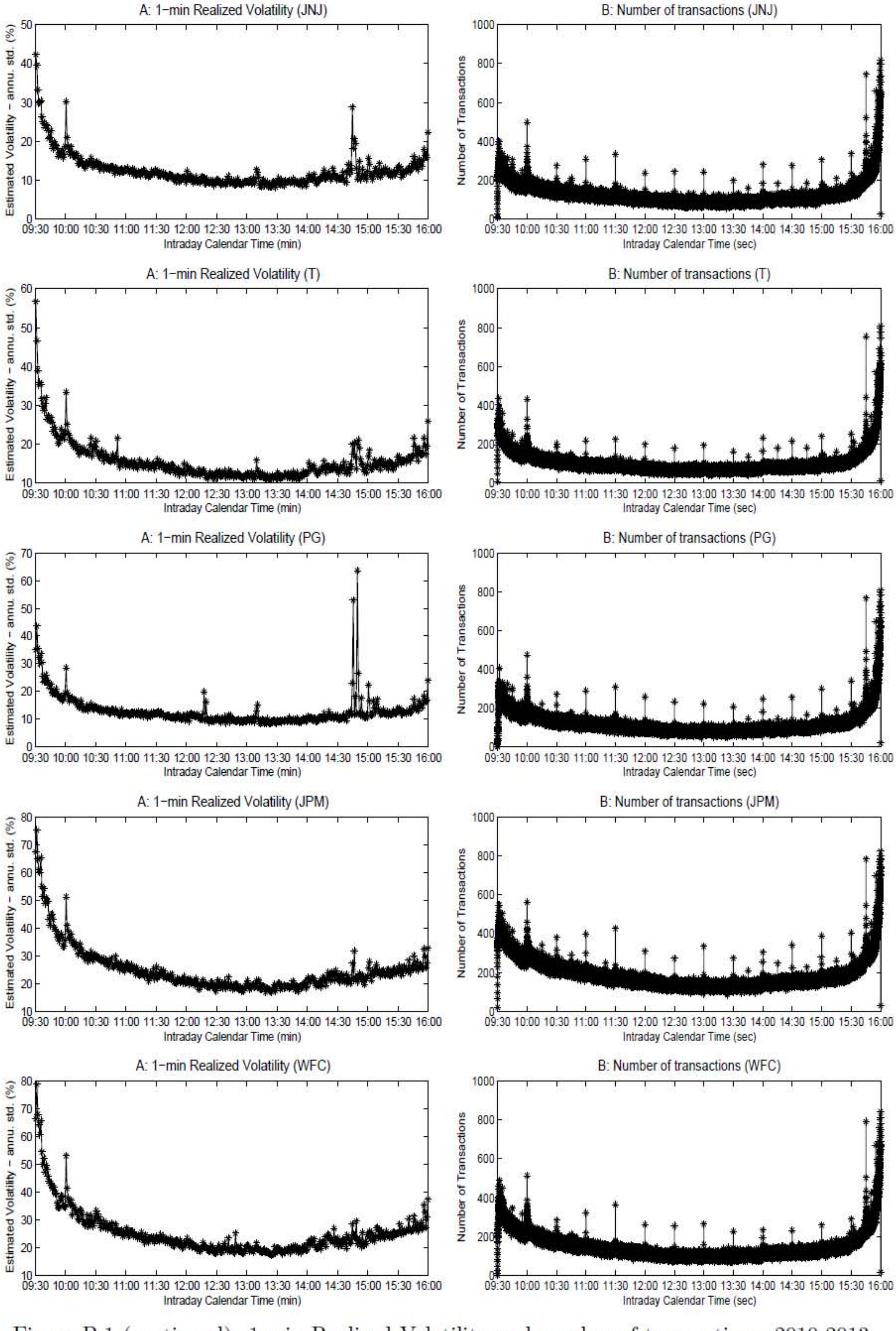

Figure B.1 (continued): 1-min Realized Volatility and number of transactions, 2010-2013.

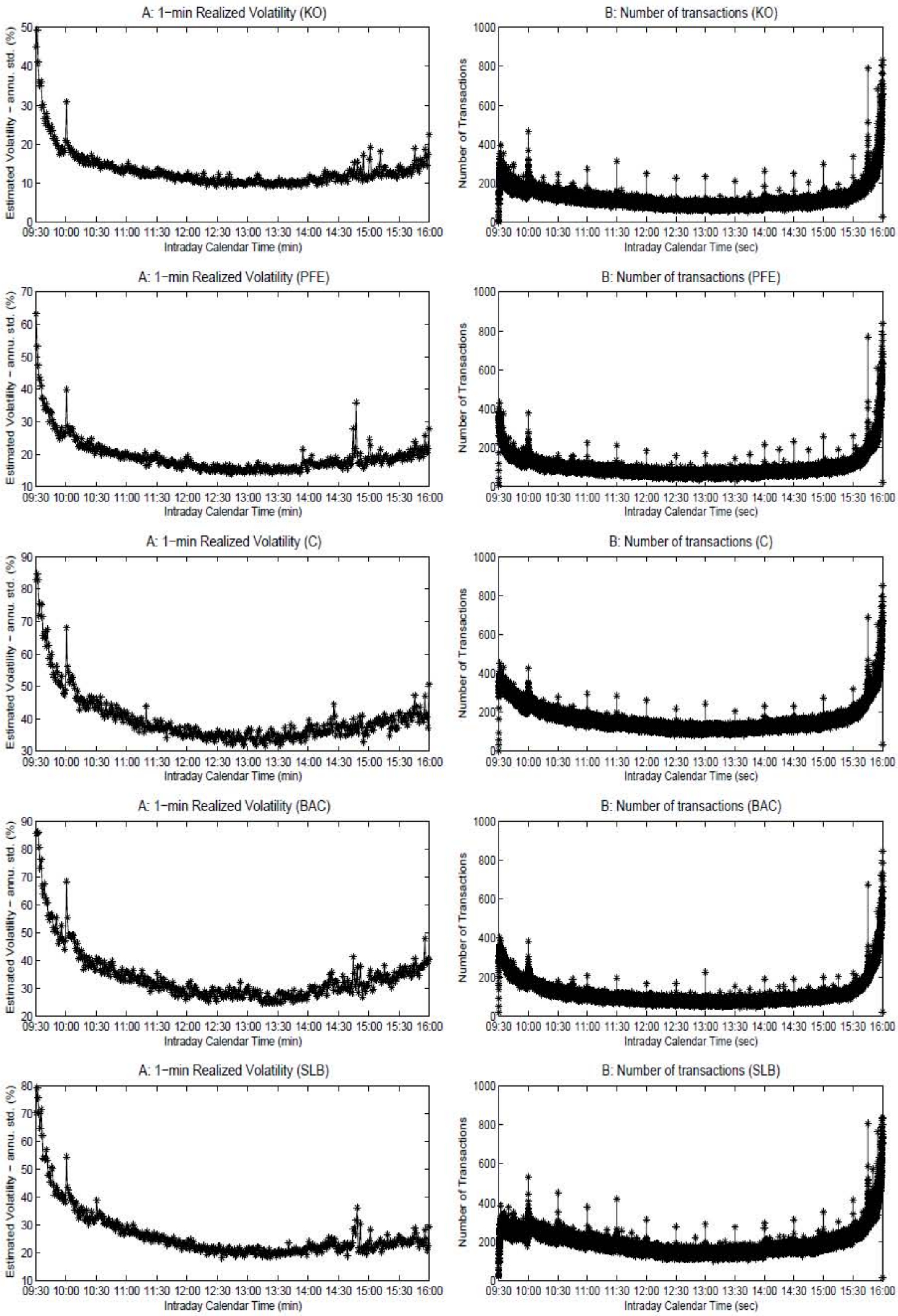

Figure B.1 (continued): 1-min Realized Volatility and number of transactions, 2010-2013.

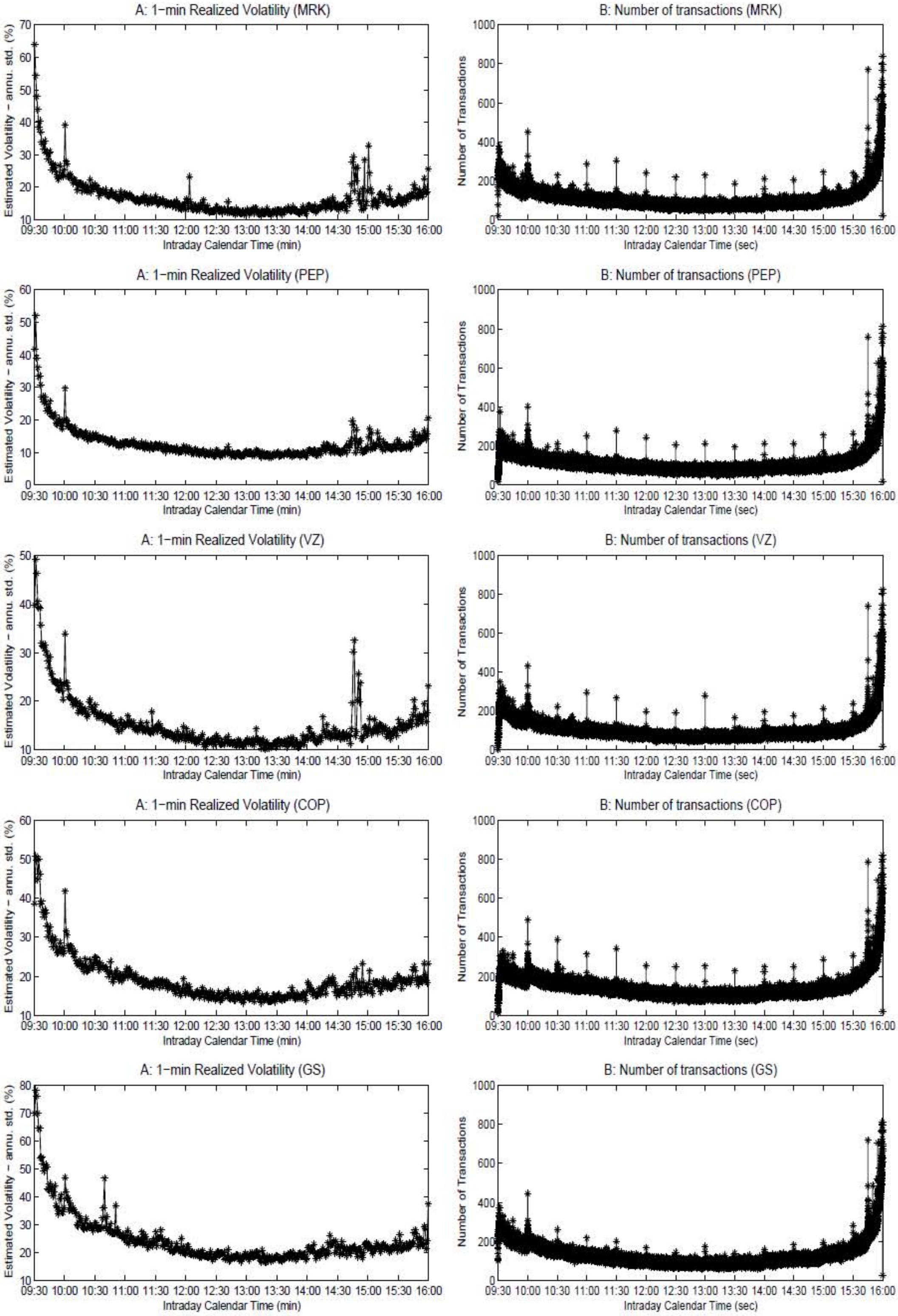

Figure B.1 (continued): 1-min Realized Volatility and number of transactions, 2010-2013.

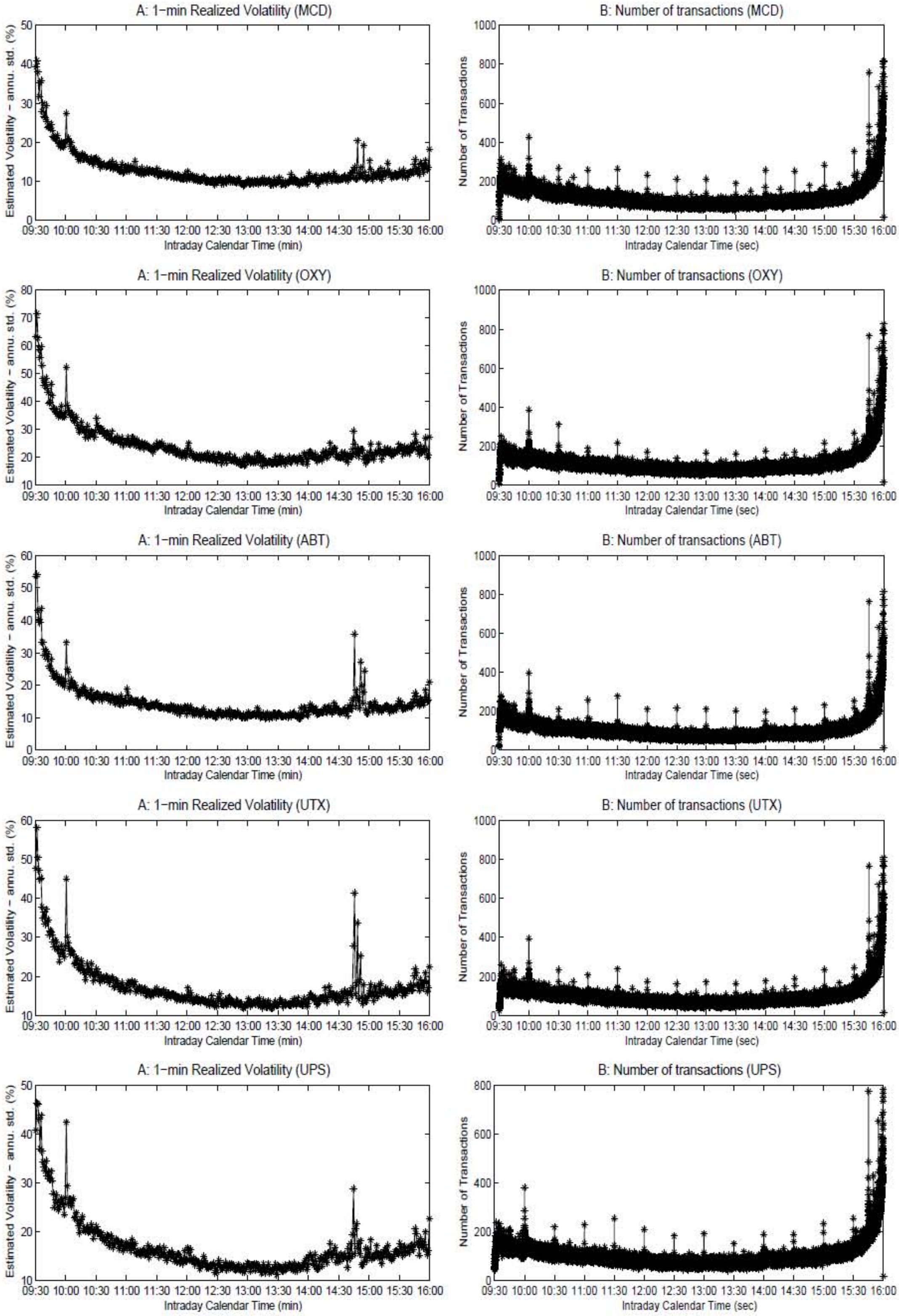

Figure B.1 (continued): 1-min Realized Volatility and number of transactions, 2010-2013.

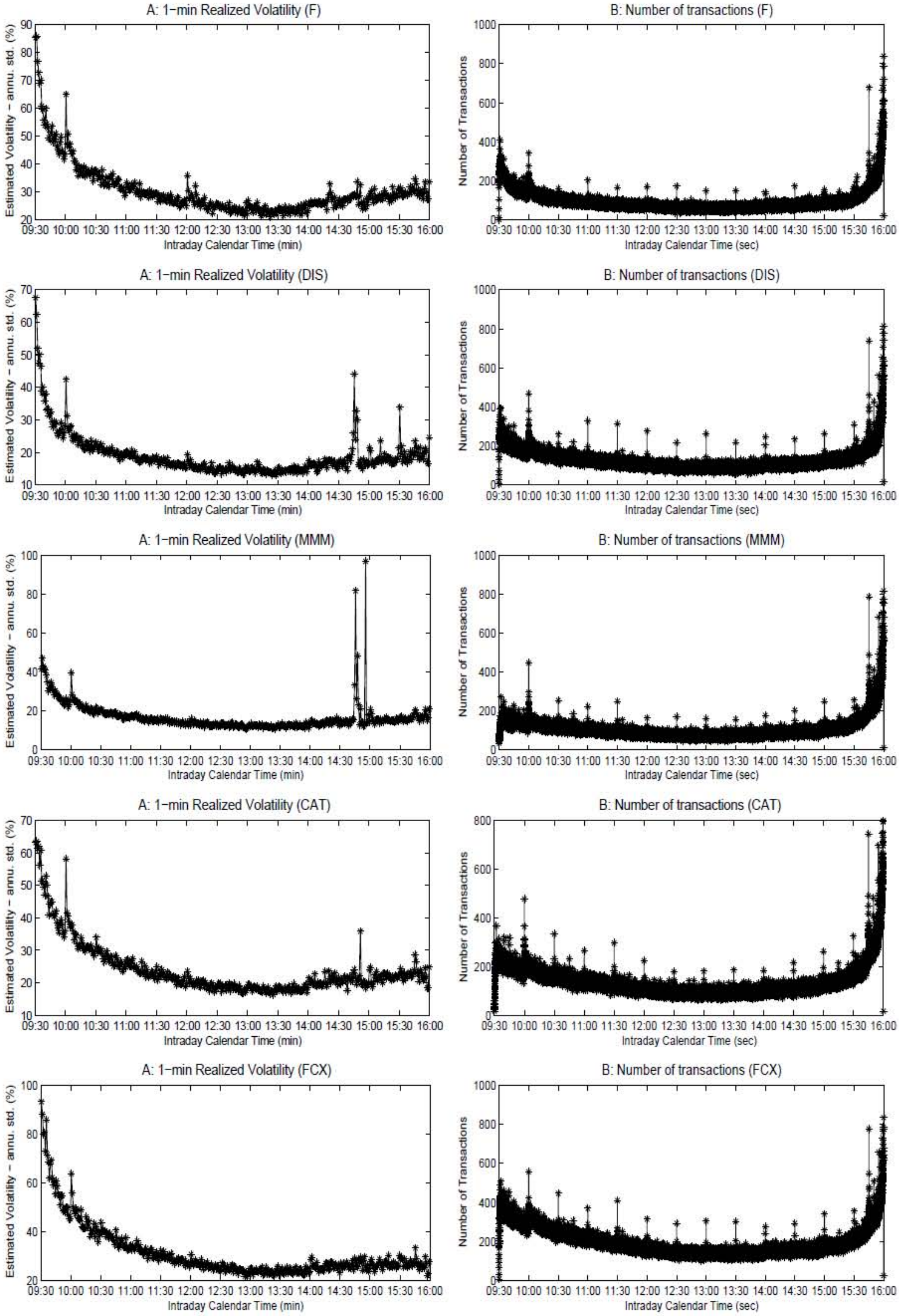

Figure B.1 (continued): 1-min Realized Volatility and number of transactions, 2010-2013.

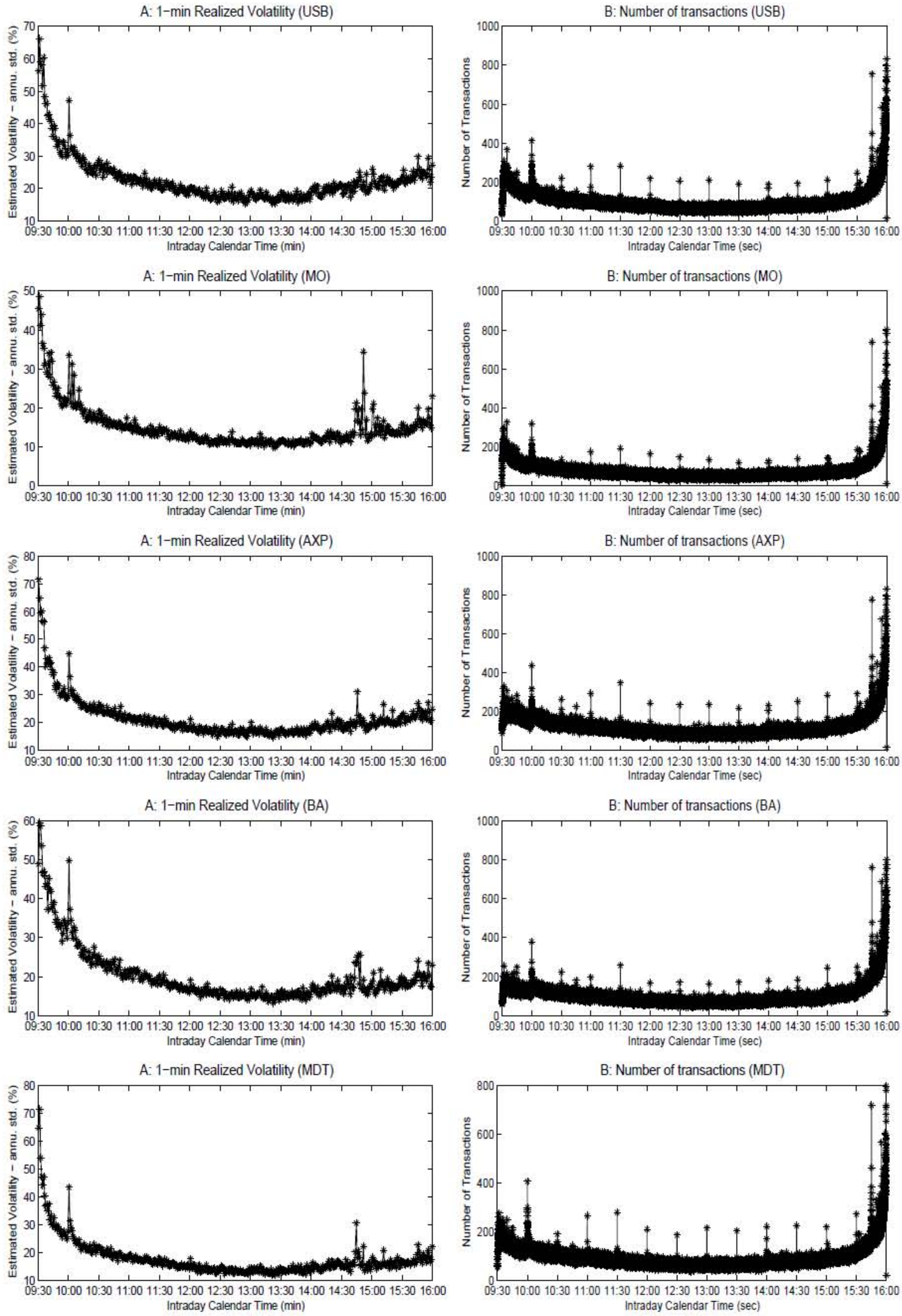

Figure B.1 (continued): 1-min Realized Volatility and number of transactions, 2010-2013.

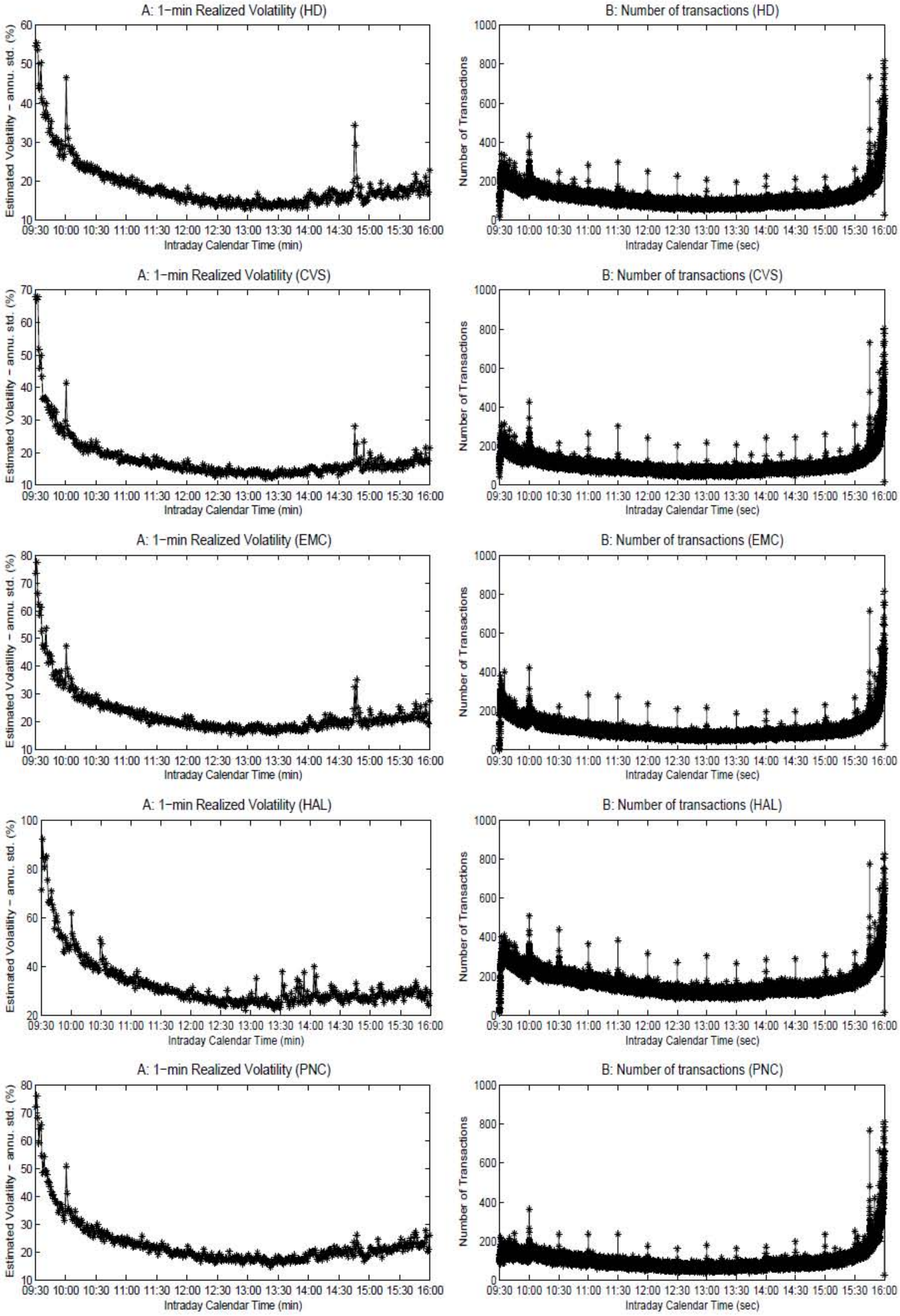

Figure B.1 (continued): 1-min Realized Volatility and number of transactions, 2010-2013.

Appendix C Appendix of Chapter 4

C.1 Asymptotic Properties of the Noise-Variance Estimates

Suppose the latent logarithmic price process $\{X_t\}$ is an Itô process following Assumption 1 with observed log-price $\{Y_t\}$. Assume the noise $\{\varepsilon_t\}$ to be iid with $\mathrm{E}(\varepsilon_t) = 0$, $\{\varepsilon_t\}$ and $\{X_t\}$ to be independent and $\mathrm{E}(\varepsilon_t^4) < \infty$. Denote the full grid as $\mathscr{G} = \{t_0, t_1, \cdots, t_n\}$ under the given sampling scheme. Define

$$M = \frac{1}{\sqrt{n}} \sum_{t_i \in \mathscr{G}} \left(\varepsilon_{t_i}^2 - \mathrm{E}(\varepsilon_t^2)\right) \tag{C.1.1}$$

and

$$N_K = \frac{1}{\sqrt{n}} \sum_{t_i \in \mathscr{G}_K^k, k=1,\cdots,K} \varepsilon_{t_i} \varepsilon_{t_{i-K}}. \tag{C.1.2}$$

Following the proof of Theorem A.1 in Zhang *et al.* (2005), (M, N_K) are the end points of martingales with respect to filtration $\mathscr{F}_i = \sigma(\varepsilon_{t_j}, j \le i, X_t, \text{ all } t)$ and the (discrete-time) predictable quadratic variations are

$$\langle M,\, M\rangle = \frac{1}{n} \sum_{t_i \in \mathscr{G}} \mathrm{Var}\left(\varepsilon_{t_i}^2 - \mathrm{E}(\varepsilon_t^2) | \mathscr{F}_{t_i - 1}\right) = \mathrm{Var}(\varepsilon_t^2), \tag{C.1.3}$$

$$\langle N_K,\, N_K\rangle = \frac{1}{n} \sum_{t_i \in \mathscr{G}_K^k, k=1,\cdots,K} \mathrm{Var}\left(\varepsilon_{t_i} \varepsilon_{t_{i-K}} | \mathscr{F}_{t_i - 1}\right) = (\mathrm{E}(\varepsilon_t^2))^2 + o_p(1) \tag{C.1.4}$$

and

$$\langle M,\, N_K\rangle = \frac{1}{n}\sum_{t_i\in\mathscr{G}_K^k,k=1,\cdots,K} \operatorname{Cov}\left(\varepsilon_{t_i}^2 - \mathrm{E}(\varepsilon_t^2), \varepsilon_{t_i}\varepsilon_{t_{i-K}}|\mathscr{F}_{t_i-1}\right) = o_p(1). \tag{C.1.5}$$

Based on the assumptions, the conditional Lindeberg conditions are satisfied. By the martingale central limit theorem in Hall and Heyde (1980), (M, N_K) are asymptotically independent normal with respective variance $\mathrm{Var}(\varepsilon_t^2)$ and $(\mathrm{E}(\varepsilon_t^2))^2$. Following equation (A.23) in Zhang *et al.* (2005), for fixed K, as $n \to \infty$, we have

$$\begin{aligned} J_K &\equiv \frac{1}{\sqrt{n}}\left(\sum_{t_i\in\mathscr{G}_K^k,k=1,\cdots,K}\left(Y_{t_i}-Y_{t_{i-K}}\right)^2 - \sum_{t_i\in\mathscr{G}_K^k,k=1,\cdots,K}\left(X_{t_i}-X_{t_{i-K}}\right)^2 - 2K\overline{m}_K\mathrm{E}(\varepsilon_t^2)\middle| X\right) \\ &= 2\left(M-N_K\right)+o_p(1) \\ &\xrightarrow{L} 2N(0,\mathrm{E}(\varepsilon_t^4)). \end{aligned} \tag{C.1.6}$$

Furthermore, by Theorem 3 of Zhang *et al.* (2005), we have $\left(\sum_{t_i\in\mathscr{G}_K^k}\left(X_{t_i}-X_{t_{i-K}}\right)^2 - \langle X,\, X\rangle\right) = o_p(1)$, where $\langle X,\, X\rangle$ is the quadratic variation of the true return process.

We observe that for any two different sampling frequencies K_1 and K_2, the (discrete-time) predictable quadratic covariation of N_{K_1} and N_{K_2} is

$$\begin{aligned} \langle N_{K_1},\, N_{K_2}\rangle &= \frac{1}{n}\sum_{t_i\in\mathscr{G}_{K_1}^k,k=1,\cdots,K_1} \operatorname{Cov}(\varepsilon_{t_i}\varepsilon_{t_{i-K_1}}, \varepsilon_{t_i}\varepsilon_{t_{i-K_2}}|\mathscr{F}_{t_i-1}) \\ &= \frac{\mathrm{E}(\varepsilon_t^2)}{n}\sum_{t_i\in\mathscr{G}_{K_1}^k,k=1,\cdots,K_1} \varepsilon_{t_{i-K_1}}\varepsilon_{t_{i-K_2}} \\ &= o_p(1). \end{aligned} \tag{C.1.7}$$

As a result, for any $K_1 < K_2$, by the martingale central limit theorem in Hall and

Heyde (1980), we have

$$
\begin{aligned}
(J_{K_1}, J_{K_2})' &= 2(M - N_{K_1},\ \ M - N_{K_2})' + o_p(1) \\
&\xrightarrow{L} 2\,\mathbf{N}\left(\mathbf{0},\ \begin{pmatrix} \mathrm{E}(\varepsilon_t^4) & \mathrm{Var}(\varepsilon_t^2) \\ \mathrm{Var}(\varepsilon_t^2) & \mathrm{E}(\varepsilon_t^4) \end{pmatrix}\right).
\end{aligned} \tag{C.1.8}
$$

Proof of Proposition 1. Under the assumptions as in Proposition 1, for the noise-variance estimate M1, we have

$$
\begin{aligned}
\sqrt{n}\left(\hat{\omega}^2(K) - \mathrm{E}(\varepsilon_t^2)\middle|X\right) &= \frac{\sqrt{n}}{2K\overline{m}_K}\left(\sum_{t_i \in \mathscr{G}_K^k, k=1,\cdots,K} \left(Y_{t_i} - Y_{t_{i-K}}\right)^2 - KIV - 2K\overline{m}_K \mathrm{E}(\varepsilon_t^2)\middle|X\right) \\
&= \frac{n}{K\overline{m}_K}\left(M - N_K + o_p(1)\right) \\
&\xrightarrow{L} \mathbf{N}\left(0, \mathrm{E}(\varepsilon_t^4)\right).
\end{aligned} \tag{C.1.9}
$$

Proof of Proposition 2. Under the assumptions as in Proposition 2, for the noise-variance estimate M2, we have

$$
\begin{aligned}
&\sqrt{n}\left(\check{\omega}^2(K_1,K_2) - \mathrm{E}(\varepsilon_t^2)\middle|X\right) \\
&= \frac{\sqrt{n}}{2(\overline{m}_{K_1} - \overline{m}_{K_2})}\left(\frac{1}{K_1}\left(\sum_{t_i \in \mathscr{G}_{K_1}^k, k=1,\cdots,K_1} \left(Y_{t_i} - Y_{t_{i-K_1}}\right)^2 - 2K_1\overline{m}_{K_1}\mathrm{E}(\varepsilon_t^2)\right)\right. \\
&\quad \left. - \frac{1}{K_2}\left(\sum_{t_i \in \mathscr{G}_{K_2}^k, k=1,\cdots,K_1} \left(Y_{t_i} - Y_{t_{i-K_2}}\right)^2 - 2K_2\overline{m}_{K_2}\mathrm{E}(\varepsilon_t^2)\right)\middle|X\right) \\
&= \frac{n}{\overline{m}_{K_1} - \overline{m}_{K_2}}\left(\frac{M - N_{K_1}}{K_1} - \frac{M - N_{K_2}}{K_2} + o_p(1)\right) \\
&\xrightarrow{L} \mathbf{N}\left(0, \mathrm{E}(\varepsilon_t^4) + \frac{2K_1K_2}{(K_2 - K_1)^2}\left(\mathrm{E}(\varepsilon_t^2)\right)^2\right).
\end{aligned} \tag{C.1.10}
$$

Note that when $K_1 < K_2$, for fixed K_2, the variance of $\check{\omega}^2(K_1,K_2)$ increases as K_1 increases.

We further consider the asymptotic properties of the noise-variance estimates

M1A and M2A. Define $H_{K,L,q}$ as

$$H_{K,L,q} = \sum_{l=-L}^{L} J_{K+lq} = 2\sum_{l=-L}^{L} (M - N_{K+lq}) + o_p(1).$$

Using equation (C.1.8), we have

$$(H_{K_1,L,q}, H_{K_2,L,q})' \xrightarrow{L} 2\,\mathbf{N}\left(\mathbf{0}, \begin{pmatrix} A, & B \\ B, & A \end{pmatrix}\right). \tag{C.1.11}$$

for all $(K_1 + Lq) < (K_2 - Lq)$, where $A = (2L+1)\mathrm{E}(\varepsilon_t^4) + 2L(2L+1)\mathrm{Var}(\varepsilon_t^2)$ and $B = (2L+1)^2\mathrm{Var}(\varepsilon_t^2)$.

Proof of Proposition 3. Under the assumptions as in Proposition 3, for the noise-variance estimate M1A, we have

$$\begin{aligned}\sqrt{n}\left(\hat{\omega}^2(K,L,q) - \mathrm{E}(\varepsilon_t^2)\middle|X\right) &= \frac{n}{2K\overline{m}_K(2L+1)}\left(H_{K,L,q} + o_p(1)\right) \\ &\xrightarrow{L} \mathbf{N}\left(0, \mathrm{E}(\varepsilon_t^4) - \frac{2L}{2L+1}\left(\mathrm{E}(\varepsilon_t^2)\right)^2\right).\end{aligned} \tag{C.1.12}$$

Proof of Proposition 4. Under the assumptions as in Proposition 4, for the noise-variance estimate M2A, we have

$$\begin{aligned}&\sqrt{n}\left(\breve{\omega}^2(K_1,K_2,L,q) - \mathrm{E}(\varepsilon_t^2)\middle|X\right) \\ &= \frac{n}{2(\overline{m}_{K_1} - \overline{m}_{K_2})}\left(\frac{H_{K_1,L,q}}{K_1(2L+1)} - \frac{H_{K_2,L,q}}{K_2(2L+1)} + o_p(1)\right) \\ &\xrightarrow{L} \mathbf{N}\left(0, \mathrm{E}(\varepsilon_t^4) + \frac{1}{2L+1}\left(-2L + \frac{2K_1K_2}{(K_2-K_1)^2}\right)\left(\mathrm{E}(\varepsilon_t^2)\right)^2\right).\end{aligned} \tag{C.1.13}$$

Comparing equation (C.1.13) with equation (C.1.10), we observe that M2A has smaller error variance compared with M2. Specifically, for the M2A estimates, when we use 1-min versus 2-min sampling frequencies and 1-min versus 3-min sampling frequencies, we have

$$\sqrt{n}\left(\breve{\omega}^2(K_1, 2K_1, 2, q) - \mathrm{E}(\varepsilon_t^2)\right) \xrightarrow{L} \mathbf{N}\left(0, \mathrm{E}(\varepsilon_t^4)\right) \tag{C.1.14}$$

and

$$\sqrt{n}\left(\breve{\omega}^2(K_1,3K_1,2,q)-\mathrm{E}(\varepsilon_t^2)\right)\xrightarrow{L}\mathbf{N}\left(0,\mathrm{E}(\varepsilon_t^4)-\frac{1}{2}\mathrm{E}\left((\varepsilon_t^2)\right)^2\right). \tag{C.1.15}$$

C.2 Supplementary Material

A Volatility Signature Plot

For transaction data of 40 NYSE stocks from January 2010 to April 2013, we plot the $\overline{RV}_K$ volatility signature plots under the Tick Time Sampling (TTS) scheme with sampling frequency ranges from 5-sec to 30-min in Figure C.1 and plot the $\overline{RV}_K$ and $\overline{RV}_{KA}$ volatility signature plots under the TTS scheme with sampling frequency ranges from 5-sec to 5-min in Figure C.2. Similar results under the Calendar Time Sampling (CTS) scheme are provided in Figure C.3 and Figure C.4 separately. Figure C.1 and Figure C.3 demonstrate the significant influence of sampling frequency to the Realized Volatility estimation and Figure C.2 and Figure C.4 demonstrate the complicate noise properties when transactions are sampled at high sampling frequency. Different stocks tend to exhibit different patterns and these figures suggest that the volatility signature plot is a quite simple but helpful and powerful tool in high-frequency financial data analysis. We plot the $\overline{RV}_K$ and $\overline{RV}_{KA}$ volatility signature plots for the simulated transaction data in Figure C.5, where the sparsity parameter are equal to 5-sec, 10-sec and 20-sec.

B Noise-to-signal Ratio (NSR) Estimation

Table C.1-C.4 report the mean error (ME) and root mean-squared error (RMSE) of M1, M1A, M2, M2A and M3 using $\overline{RV}_K$ across different sampling frequencies. M2 and M2A perform better than M1, M1A and M3 by reporting smaller RMSE values. M2A again generally outperform M2 by reporting smaller RMSE values. Compared with cases when $\overline{RV}_{KA^*}$ is used, now all noise-variance estimates report much worse unbiased property and much larger RMSE values. We estimate the noise-to-signal ratio (NSR) value of 40 stocks using transaction data. Table C.5 reports the estimated NSR when we sample transactions at 30-min sampling frequency.

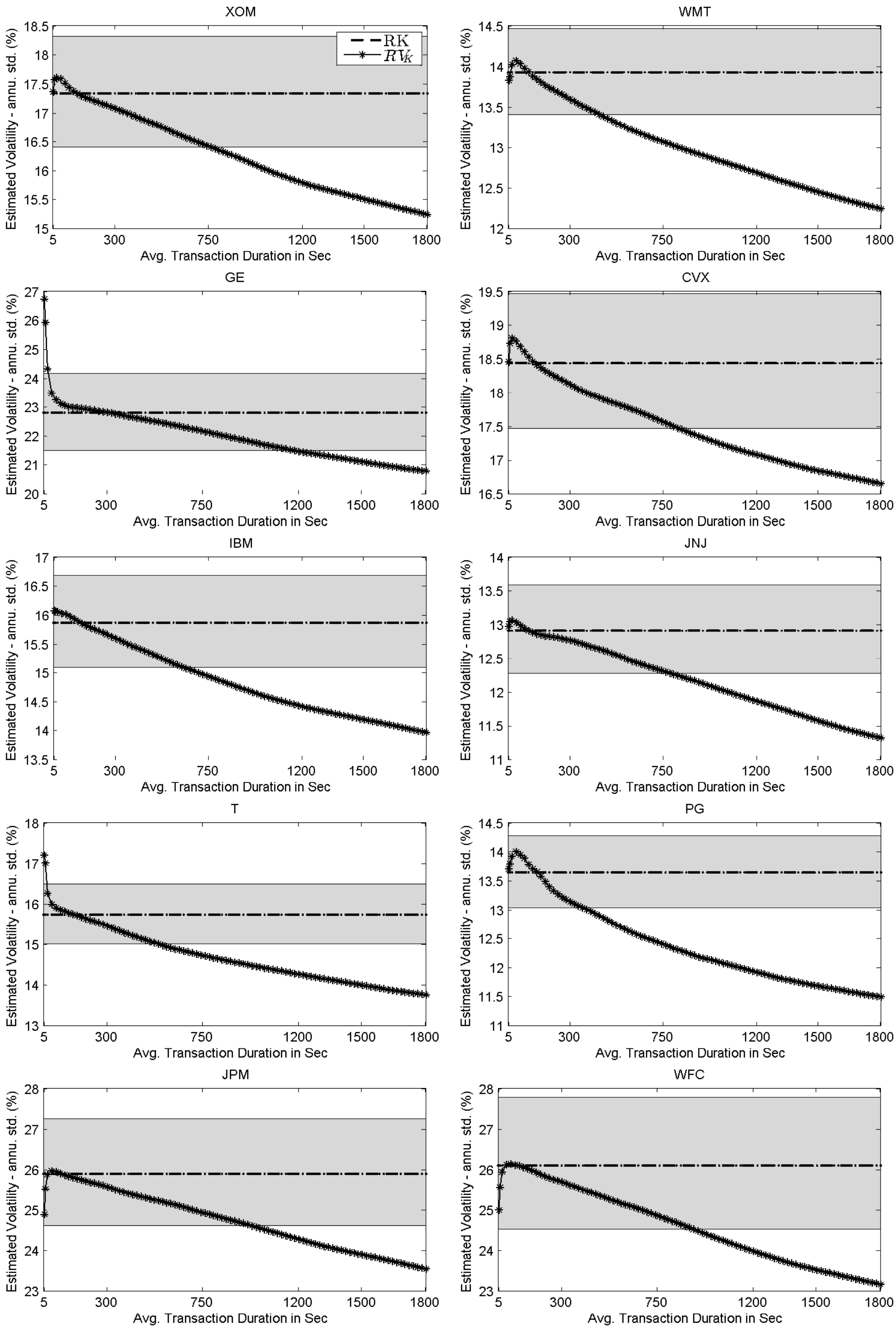

Figure C.1: $\overline{RV}_K$ volatility signature plots under the TTS scheme, 2010-2013.

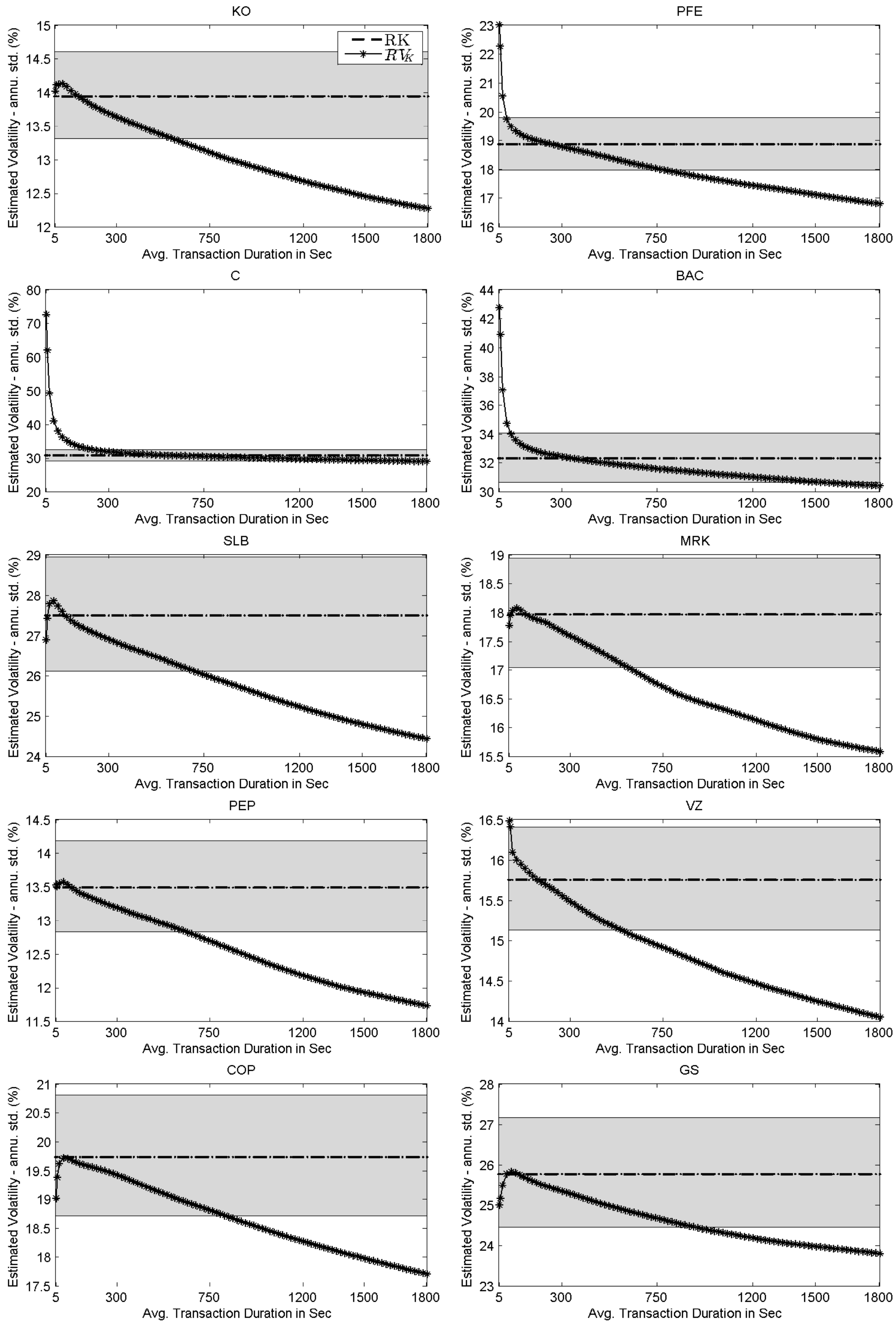

Figure C.1 (continue): $\overline{RV}_K$ volatility signature plots under the TTS scheme, 2010-2013.

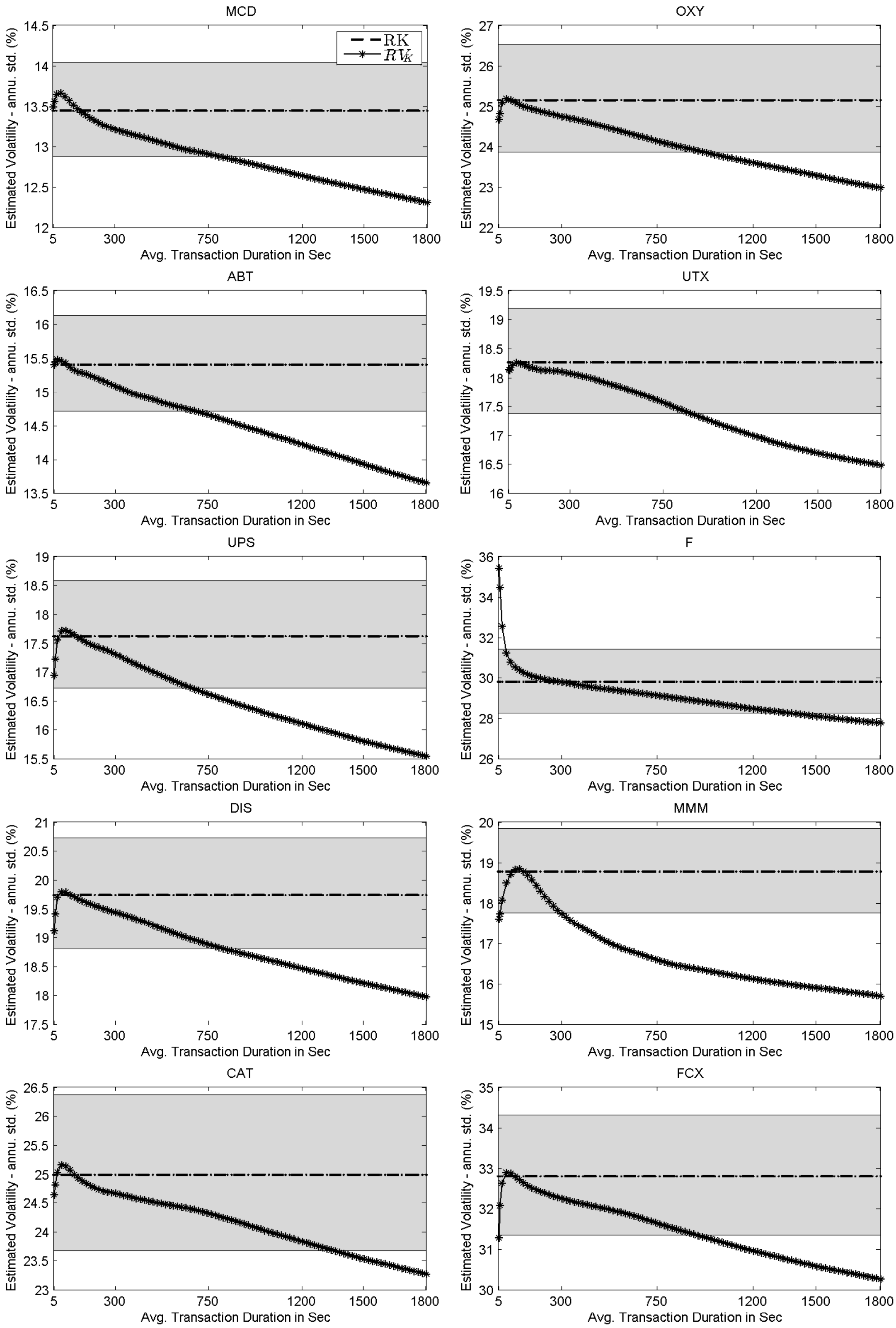

Figure C.1 (continue): $\overline{RV}_K$ volatility signature plots under the TTS scheme, 2010-2013.

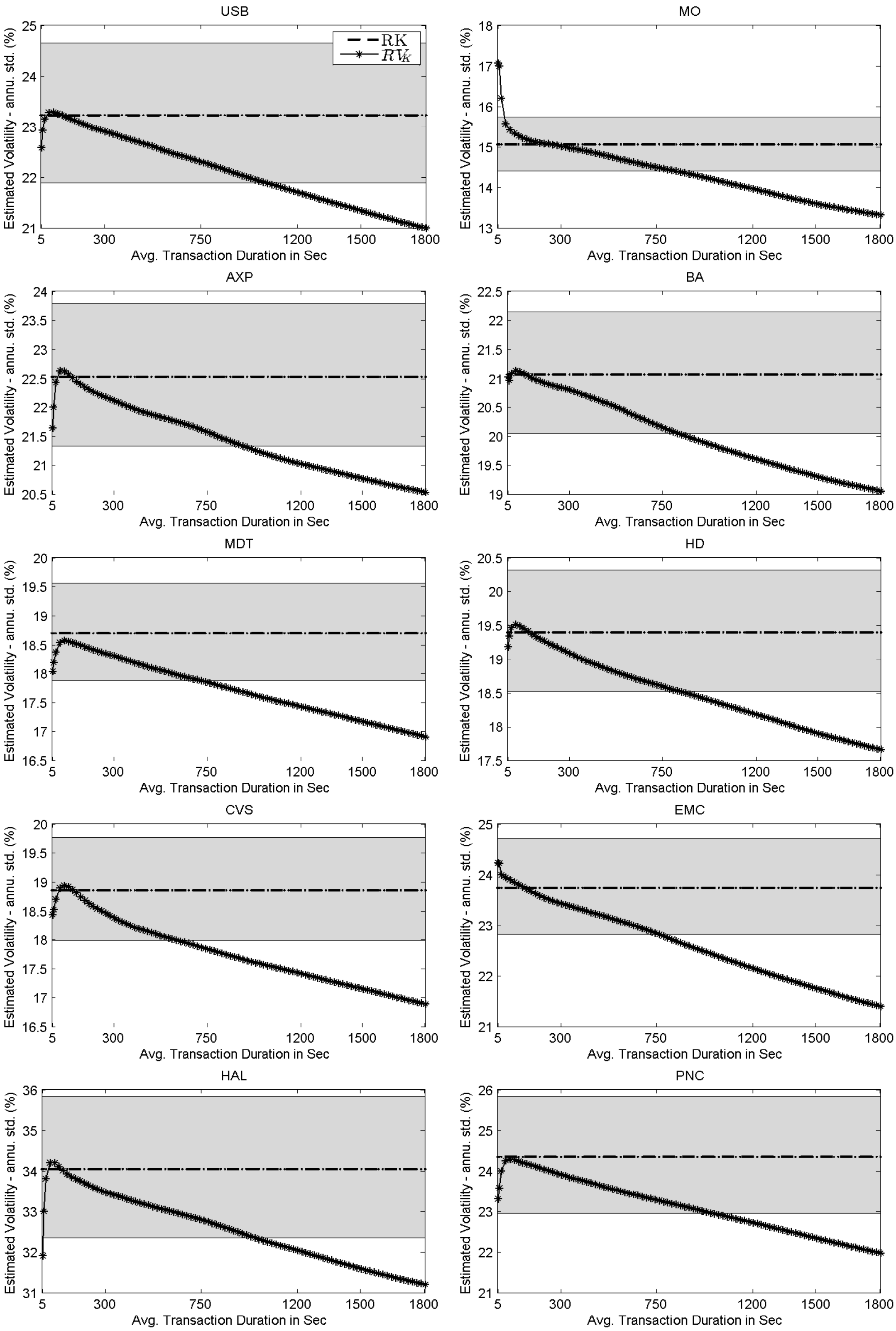

Figure C.1 (continue): $\overline{RV}_K$ volatility signature plots under the TTS scheme, 2010-2013.

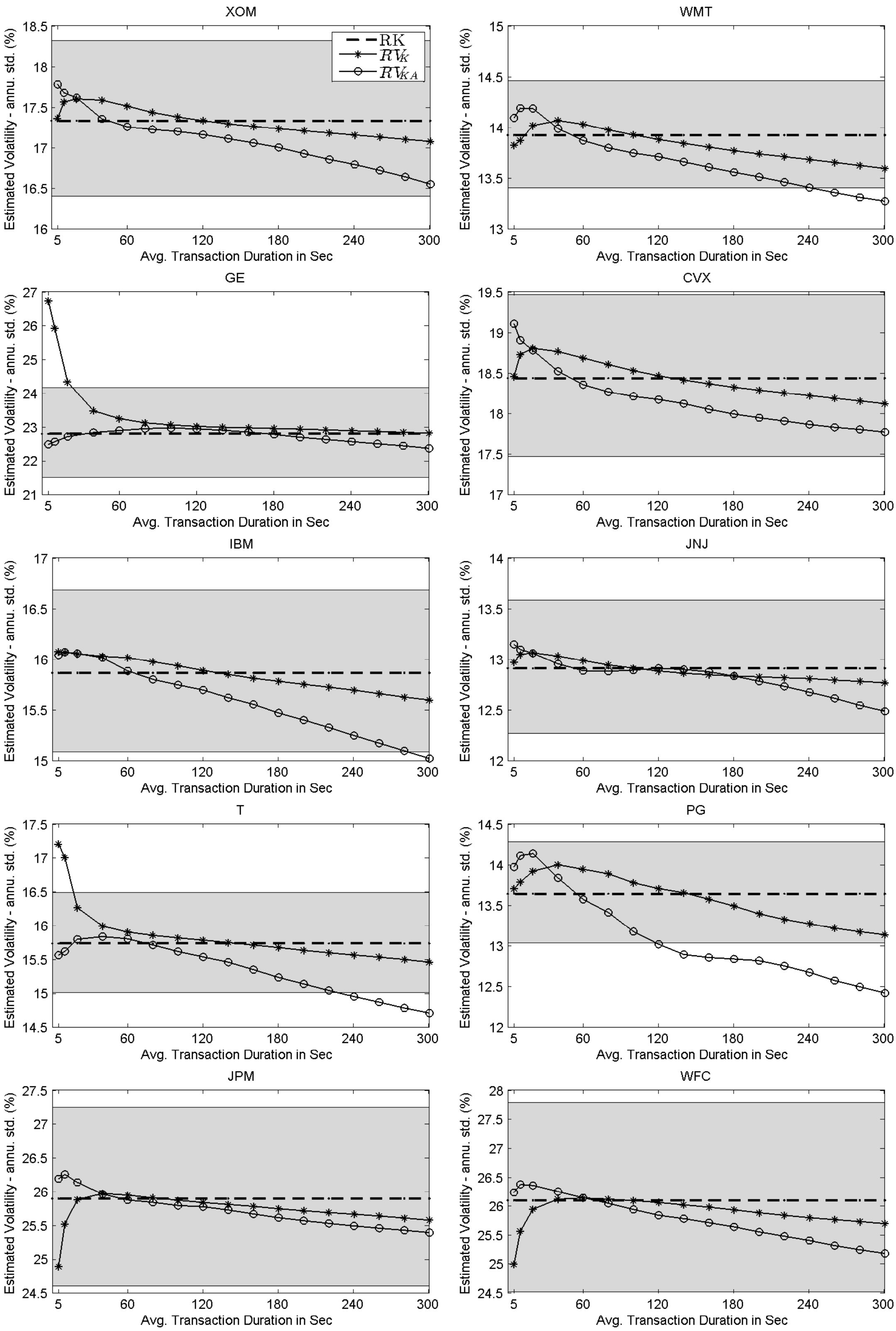

Figure C.2: $\overline{RV}_K$ and $\overline{RV}_{KA}$ volatility signature plots under the TTS scheme, 2010-2013.

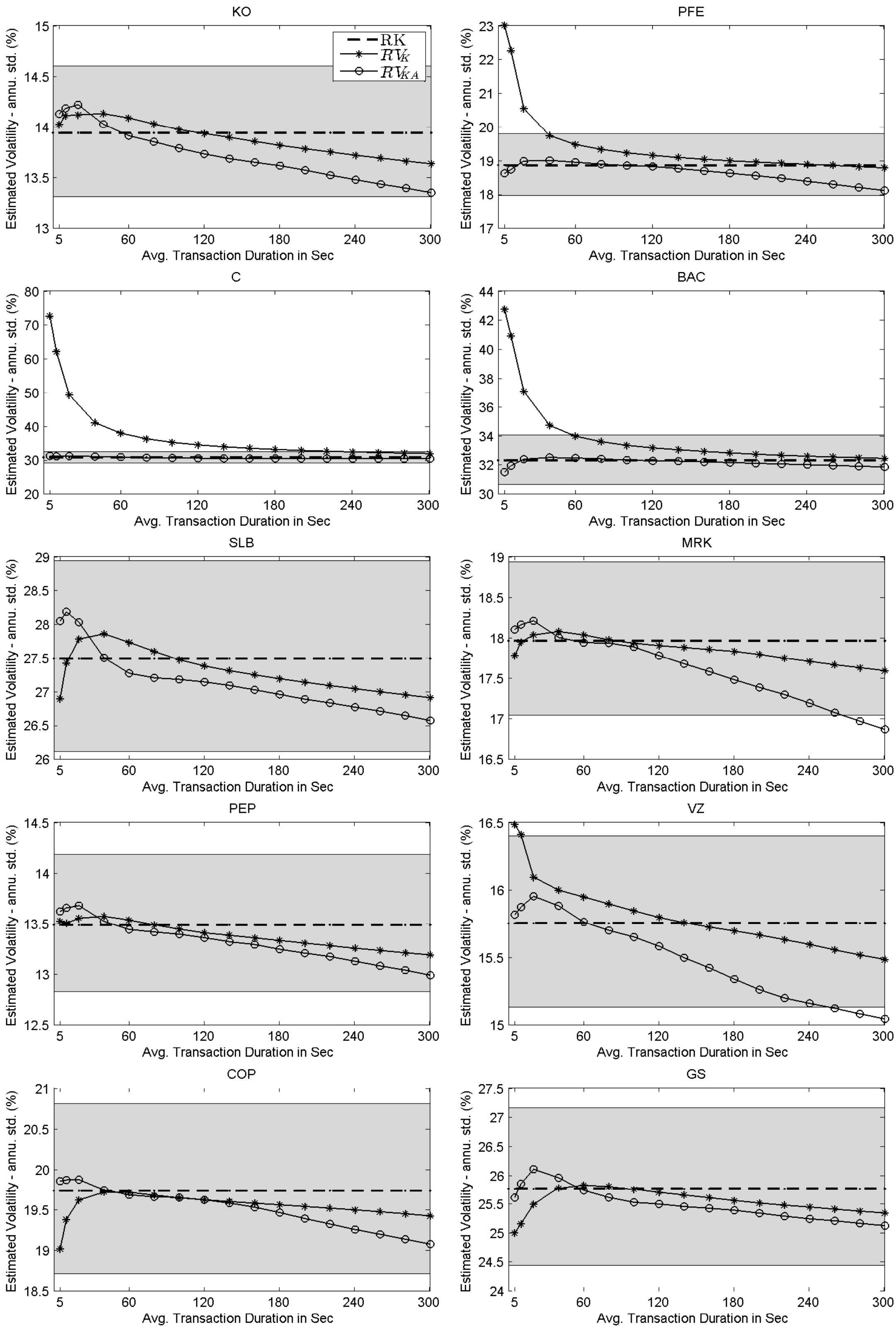

Figure C.2 (continue): $\overline{RV}_K$ and $\overline{RV}_{KA}$ volatility signature plots under the TTS scheme, 2010-2013.

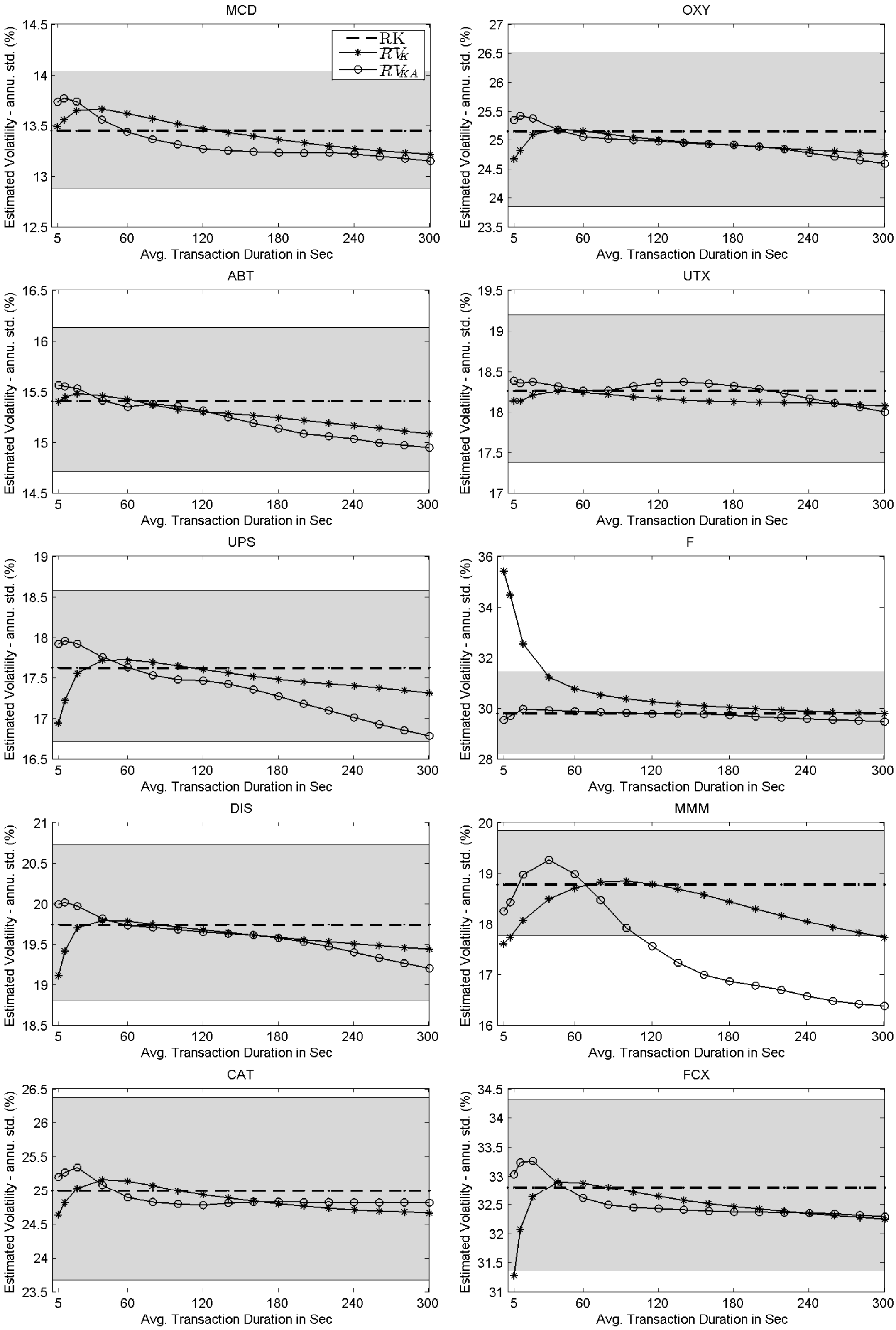

Figure C.2 (continue): $\overline{RV}_K$ and $\overline{RV}_{KA}$ volatility signature plots under the TTS scheme, 2010-2013.

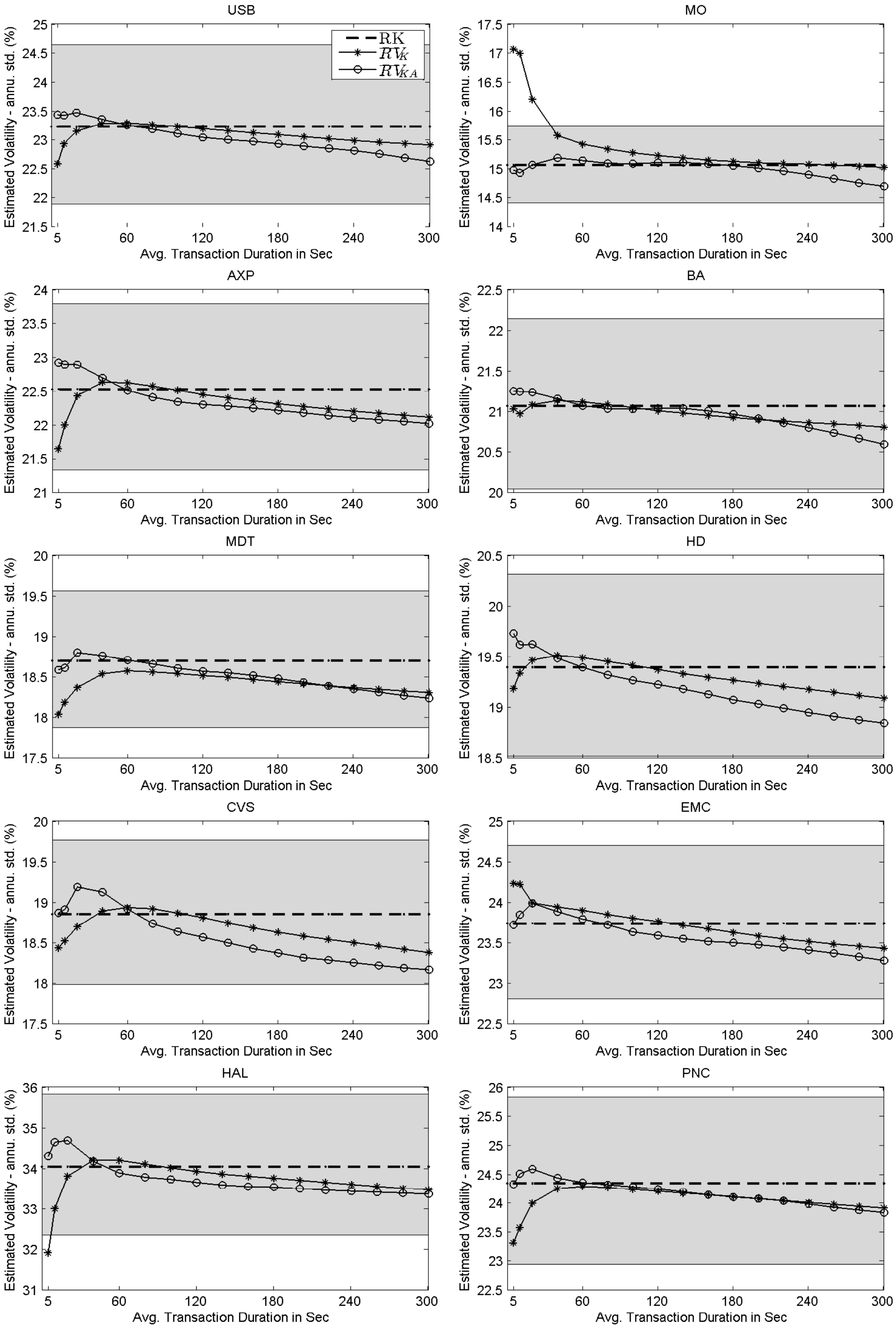

Figure C.2 (continue): $\overline{RV}_K$ and $\overline{RV}_{KA}$ volatility signature plots under the TTS scheme, 2010-2013.

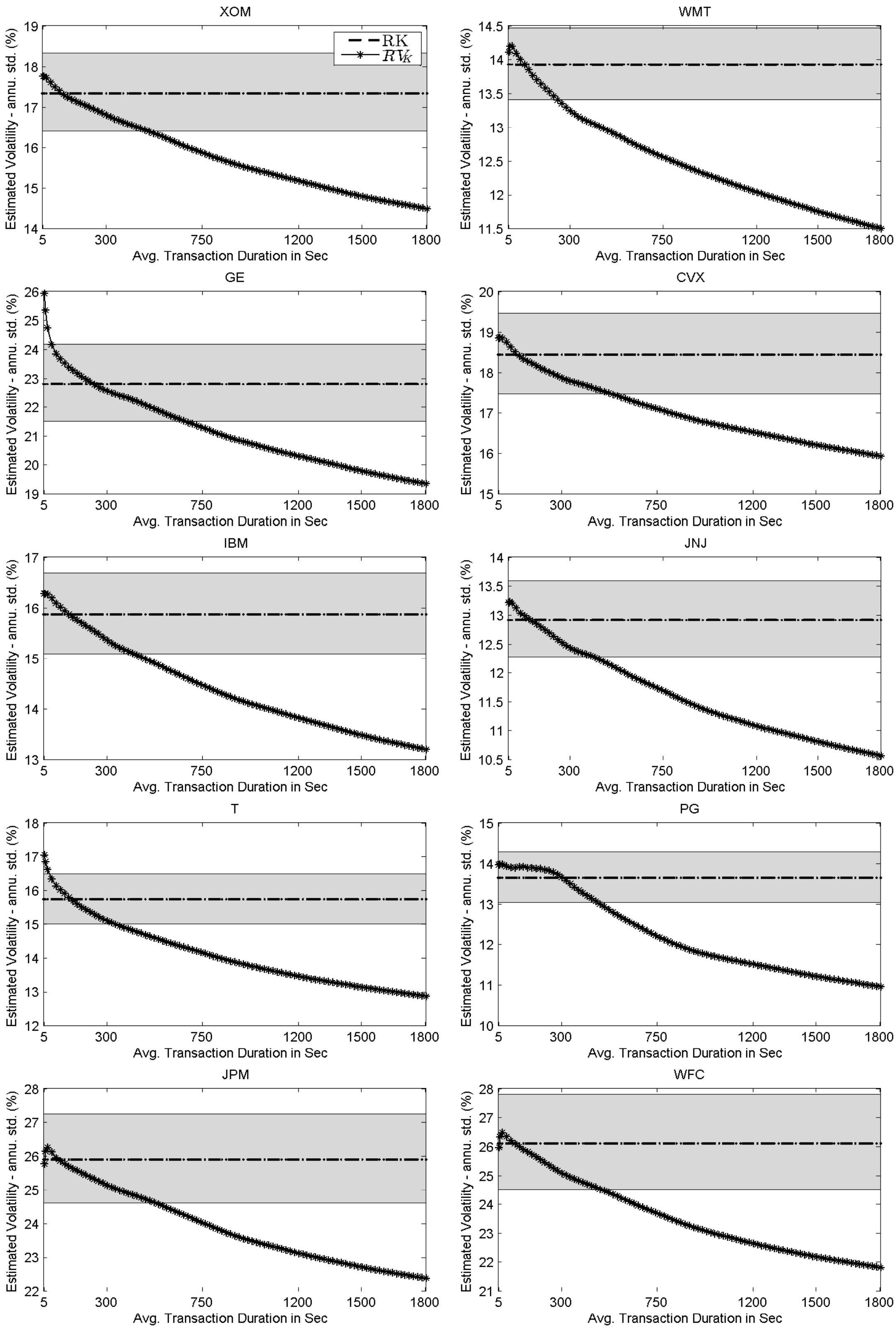

Figure C.3: $\overline{RV}_K$ volatility signature plots under the CTS scheme, 2010-2013.

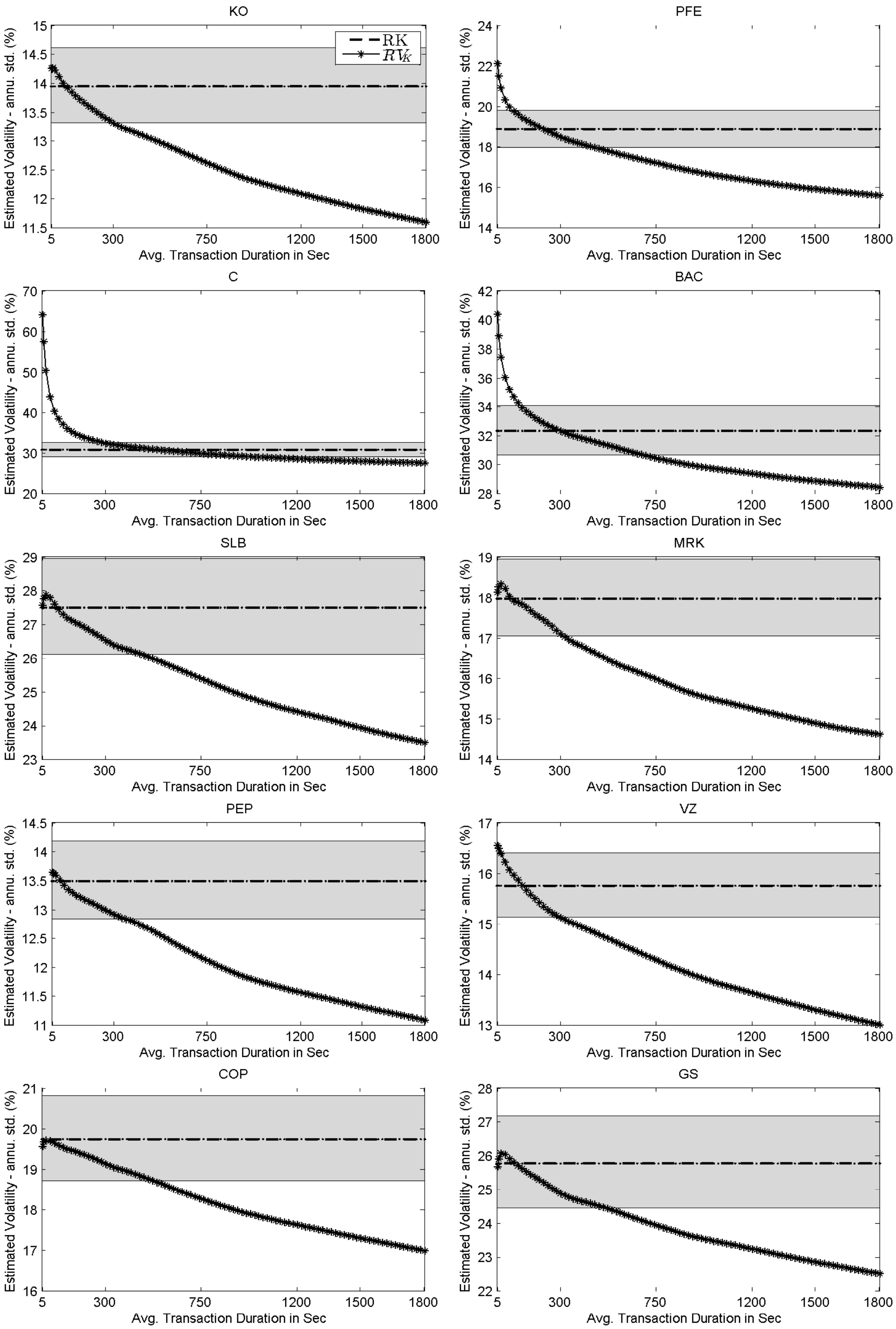

Figure C.3 (continue): $\overline{RV}_K$ volatility signature plots under the CTS scheme, 2010-2013.

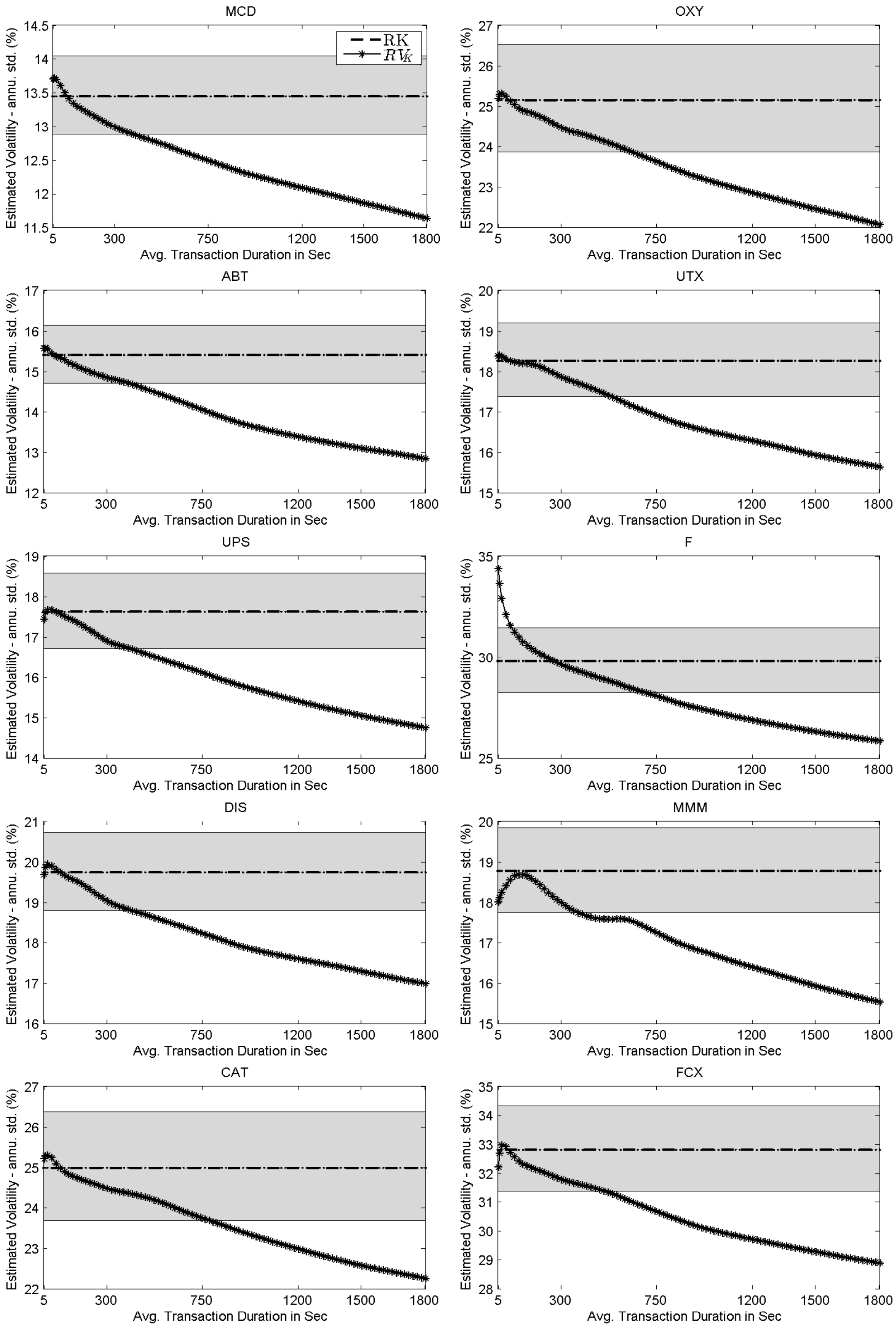

Figure C.3 (continue): $\overline{RV}_K$ volatility signature plots under the CTS scheme, 2010-2013.

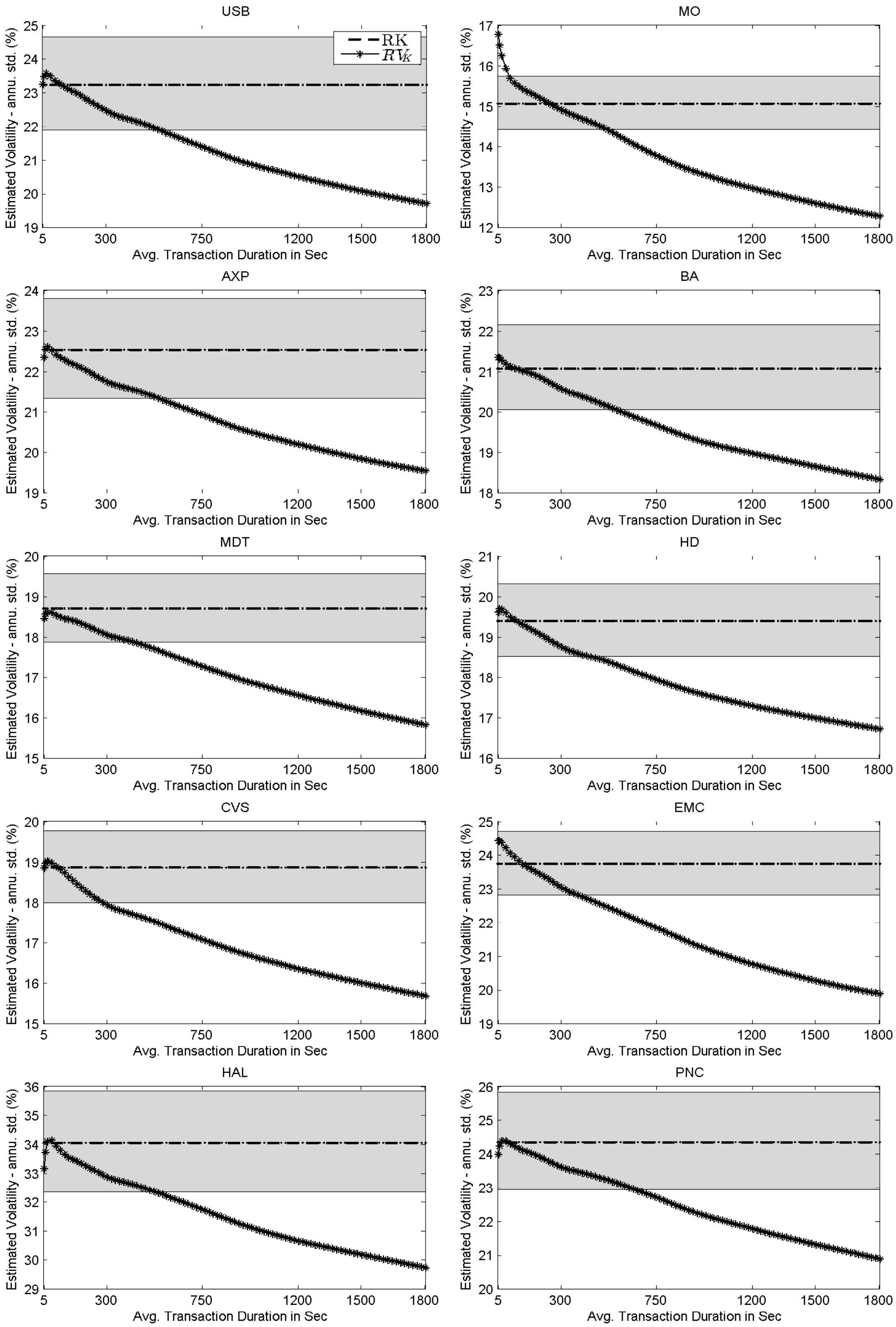

Figure C.3 (continue): $\overline{RV}_K$ volatility signature plots under the CTS scheme, 2010-2013.

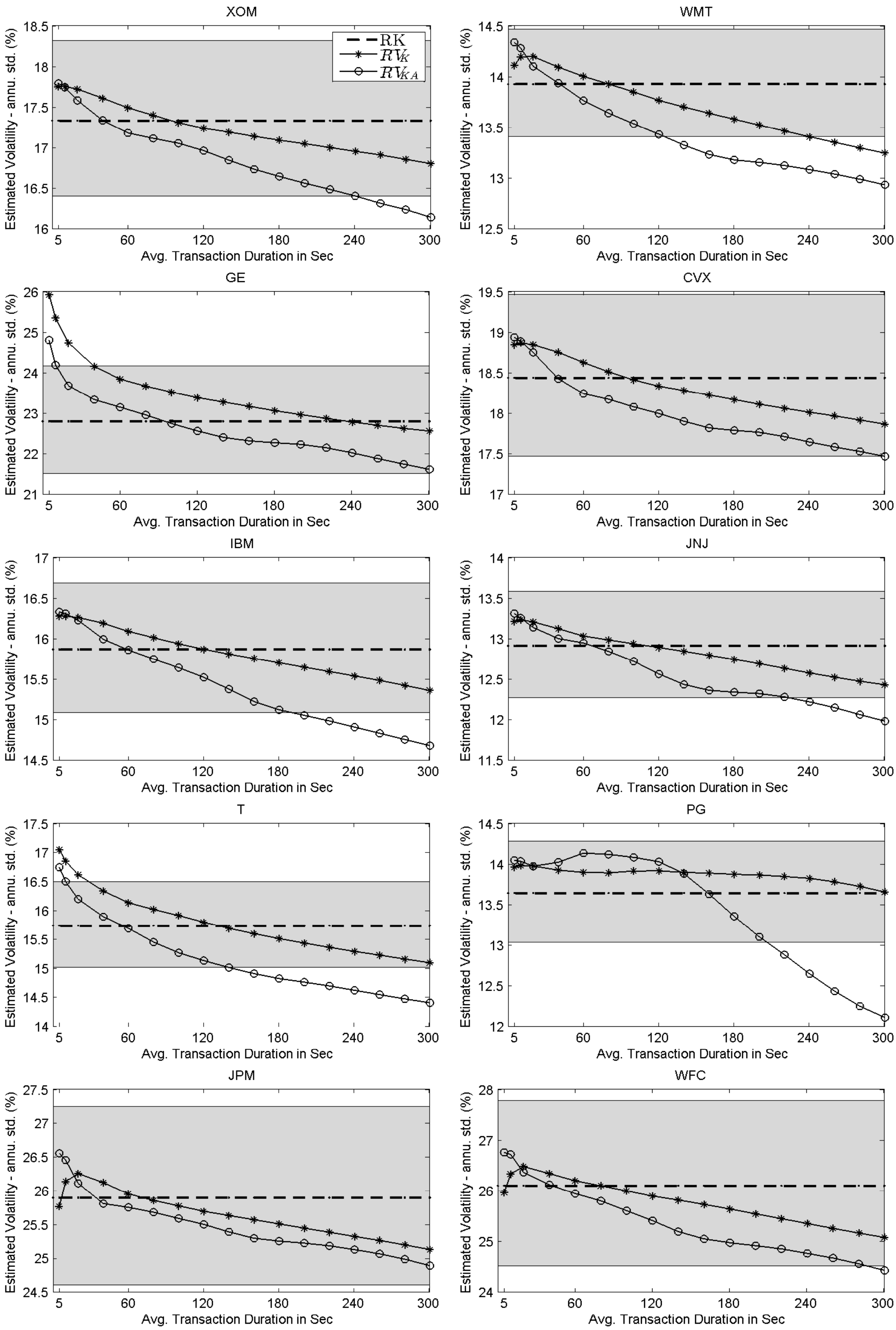

Figure C.4: $\overline{RV}_K$ and $\overline{RV}_{KA}$ volatility signature plots under the CTS scheme, 2010-2013.

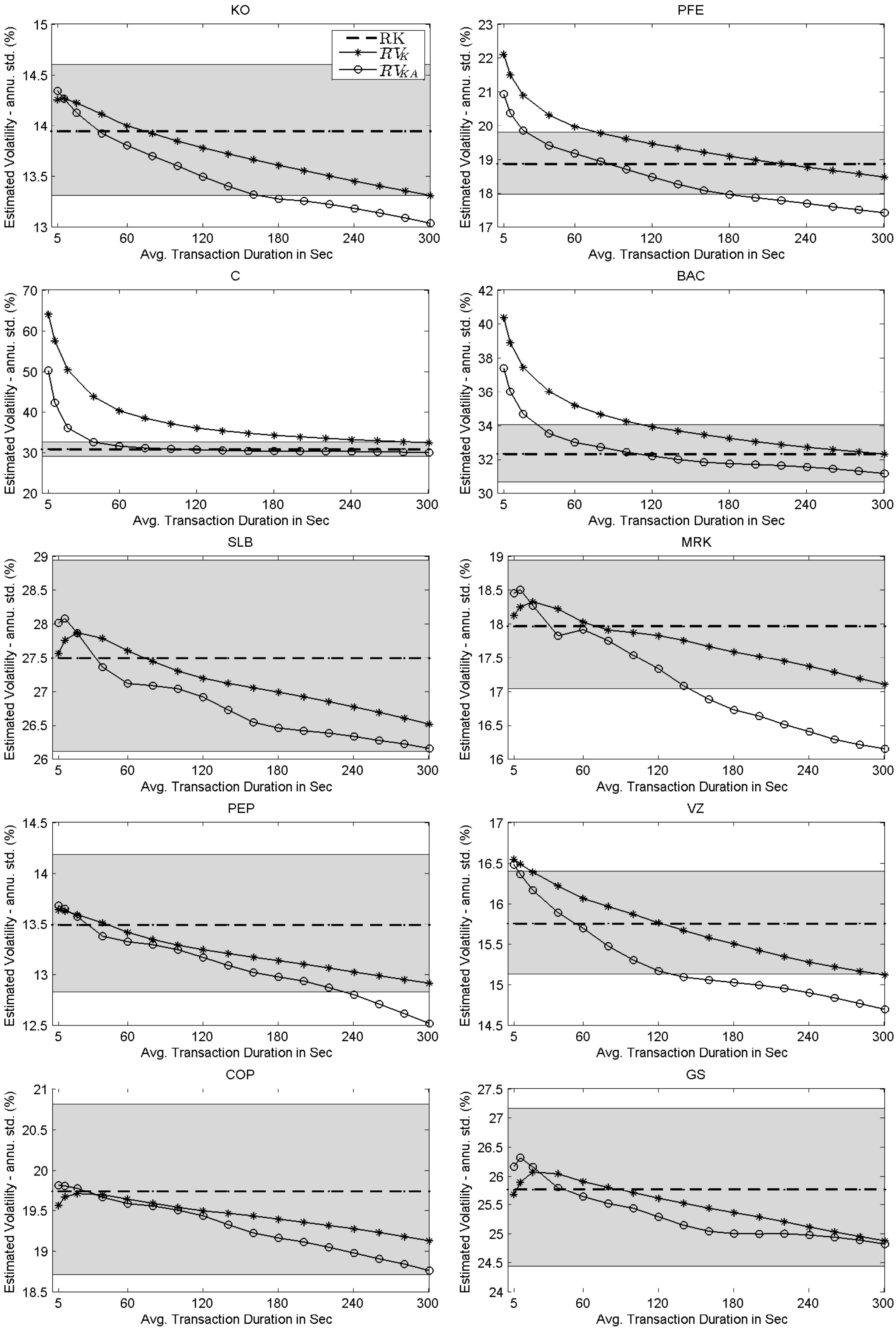

Figure C.4 (continue): $\overline{RV}_K$ and $\overline{RV}_{KA}$ volatility signature plots under the CTS scheme, 2010-2013.

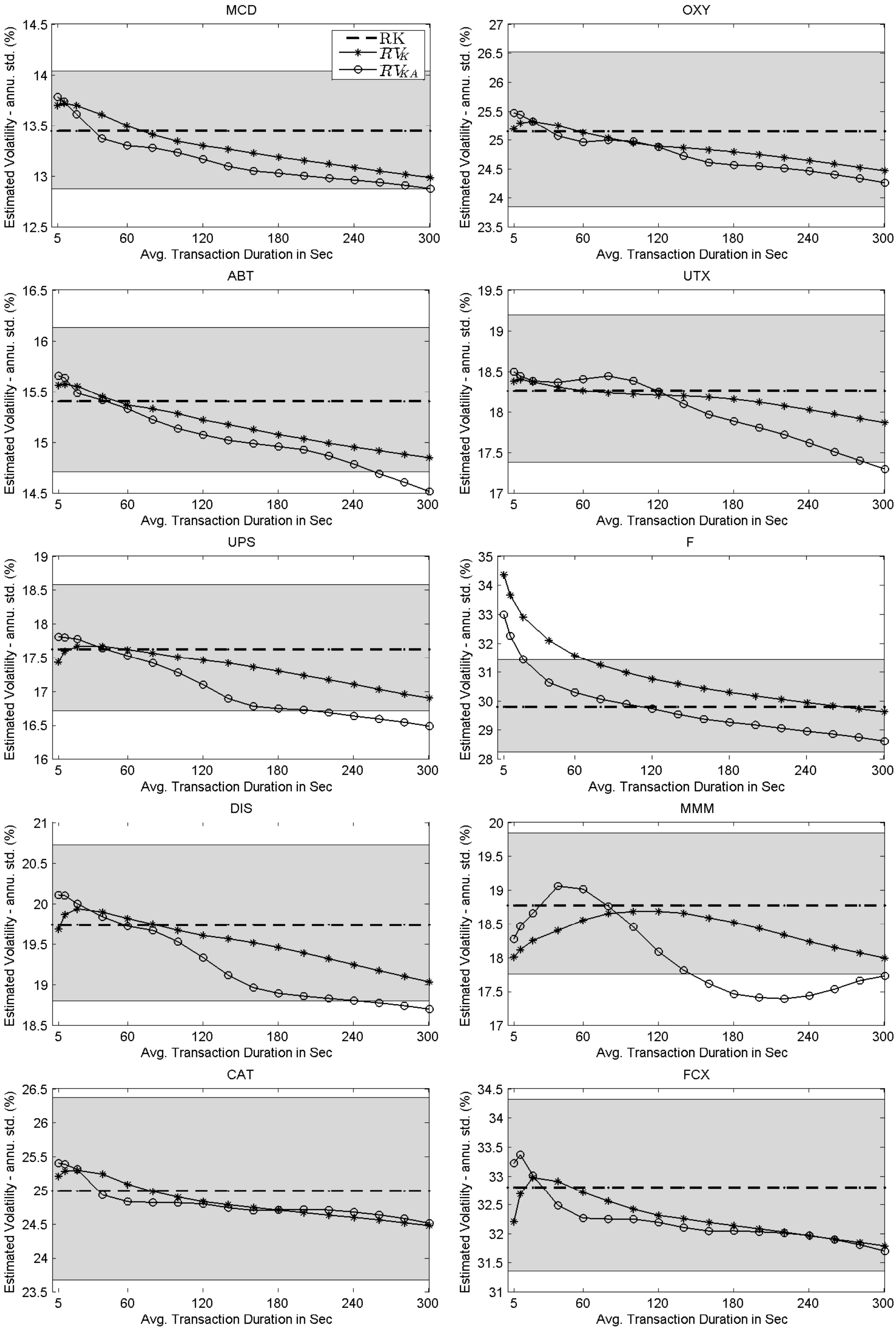

Figure C.4 (continue): $\overline{RV}_K$ and $\overline{RV}_{KA}$ volatility signature plots under the CTS scheme, 2010-2013.

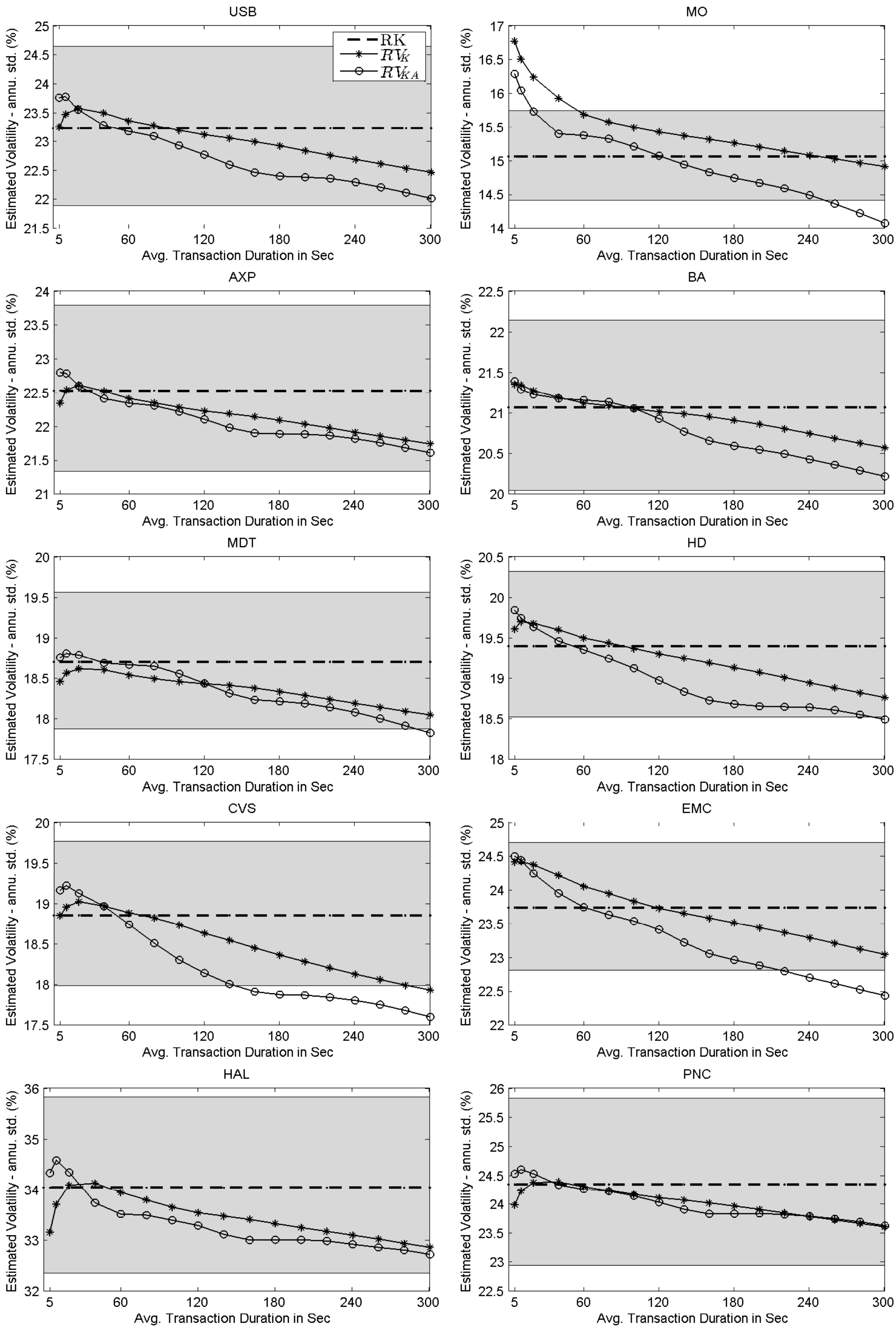

Figure C.4 (continue): $\overline{RV}_K$ and $\overline{RV}_{KA}$ volatility signature plots under the CTS scheme, 2010-2013.

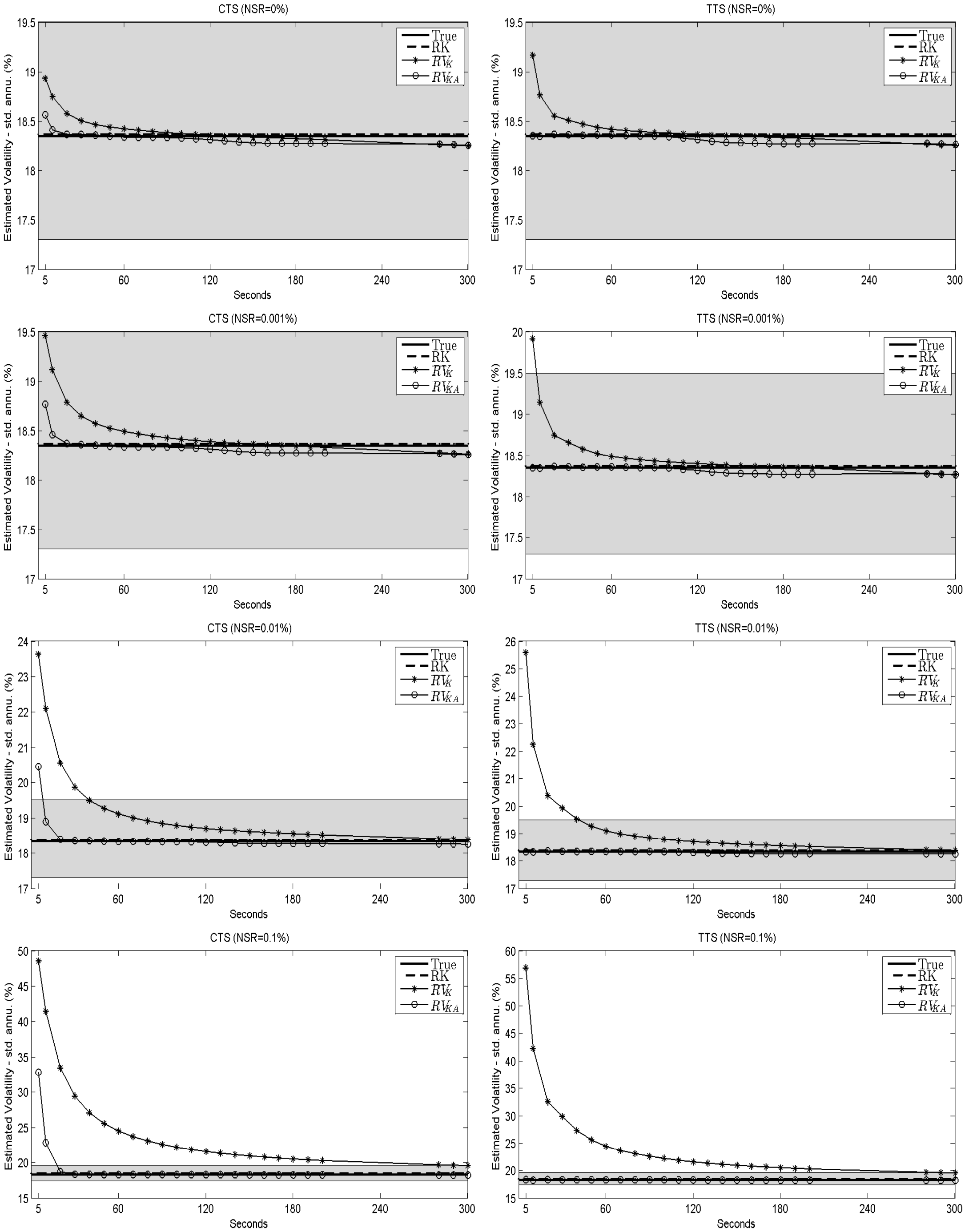

Figure C.5: Volatility signature plot of one simulation run, Sparsity=5-sec.

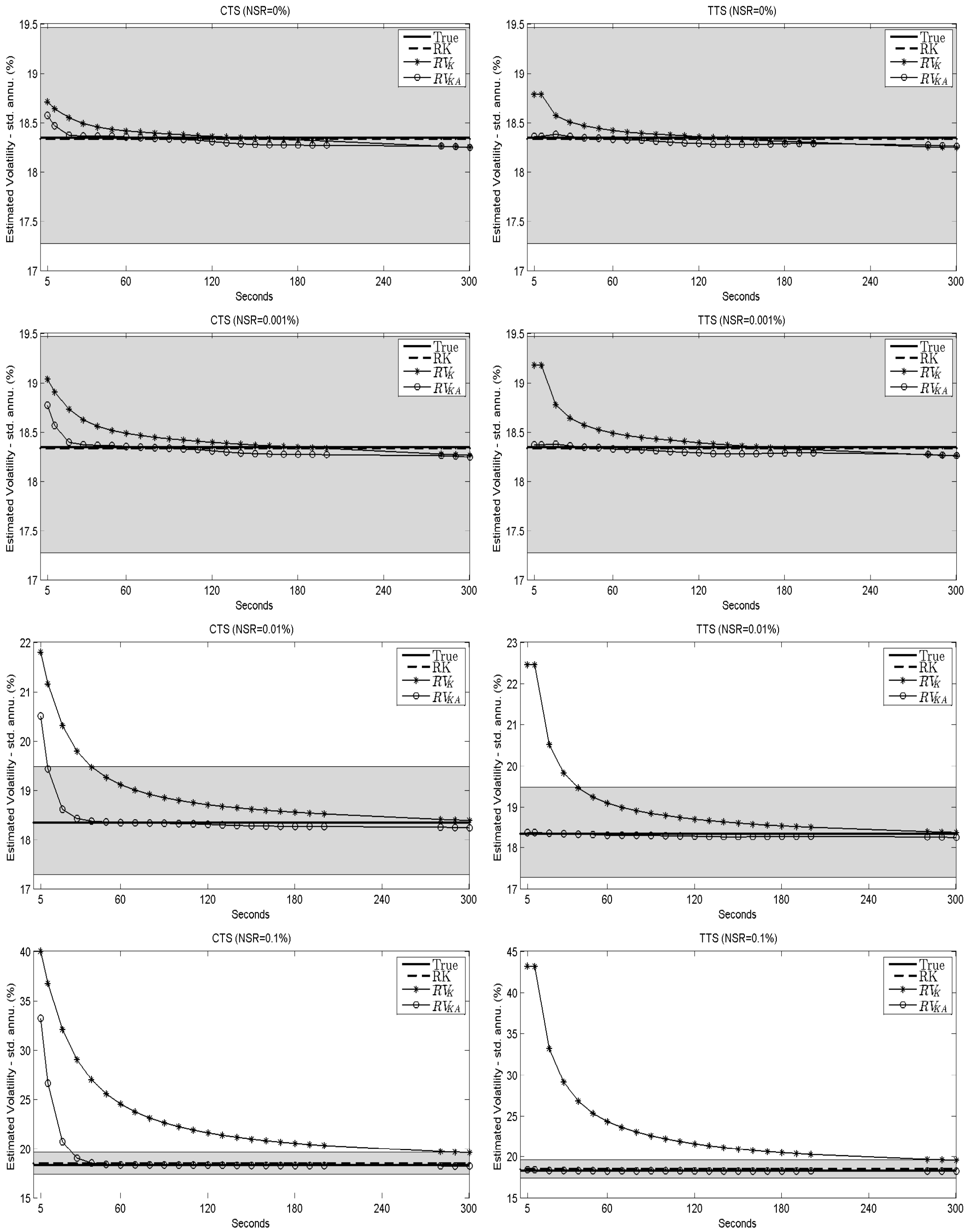

Figure C.5 (continue): Volatility signature plot of one simulation run, Sparsity=10-sec.

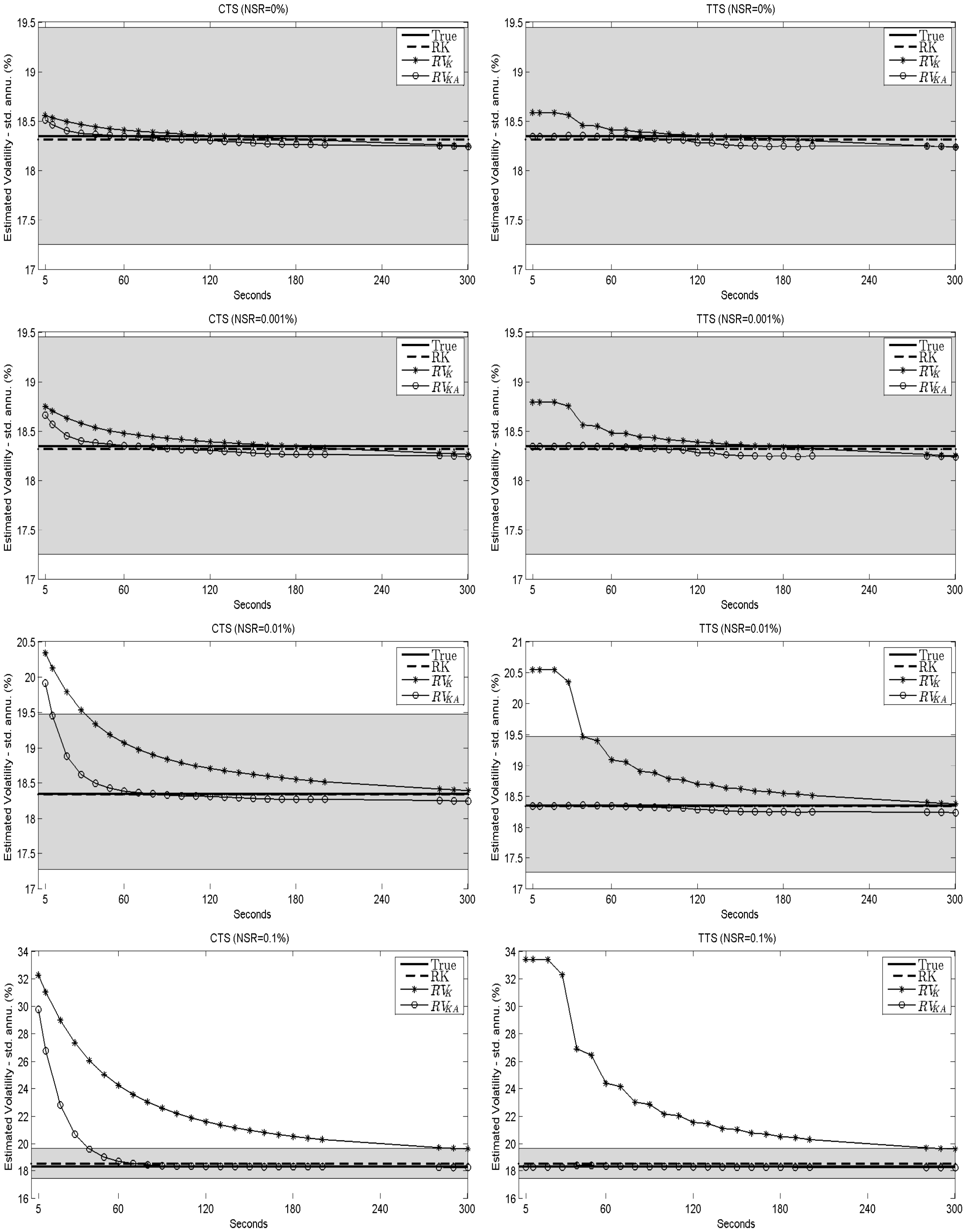

Figure C.5 (continue): Volatility signature plot of one simulation run, Sparsity=20-sec.

Table C.1: ME and RMSE for NSR estimate M1 using $\overline{RV}_K$ under the CTS and TTS schemes

Scheme	Sparsity	NSR	True NSR	ME ($\times 10^4$)					RMSE ($\times 10^4$)				
			($\times 10^4$)	1-min	3-min	5-min	10-min	30-min	1-min	3-min	5-min	10-min	30-min
CTS	5-sec	0%	0.1580	-0.0166	-0.2449	-0.7442	-3.2031	-31.7148	0.0441	0.2480	0.7545	3.2642	32.0261
		0.001%	0.2583	-0.0193	-0.2529	-0.7572	-3.2295	-31.7920	0.0453	0.2558	0.7673	3.2901	32.1024
		0.01%	1.1594	-0.0430	-0.3244	-0.8762	-3.4677	-32.5064	0.0598	0.3267	0.8847	3.5238	32.8084
		0.1%	9.9994	-0.2643	-0.9895	-1.9832	-5.6835	-39.1476	0.2683	0.9906	1.9868	5.7161	39.3885
	10-sec	0%	0.1581	-0.0164	-0.2432	-0.7431	-3.2002	-31.7214	0.0546	0.2540	0.7484	3.2466	32.0179
		0.001%	0.2584	-0.0195	-0.2523	-0.7583	-3.2302	-31.8123	0.0555	0.2627	0.7635	3.2762	32.1079
		0.01%	1.1597	-0.0486	-0.3331	-0.8932	-3.4997	-32.6178	0.0714	0.3413	0.8977	3.5419	32.9046
		0.1%	9.9805	-0.3186	-1.0793	-2.1371	-5.9877	-40.0770	0.3238	1.0825	2.1393	6.0114	40.3008
	20-sec	0%	0.1582	-0.0225	-0.2338	-0.7250	-3.1630	-31.6474	0.0680	0.2638	0.7352	3.1933	31.9126
		0.001%	0.2587	-0.0308	-0.2438	-0.7416	-3.1968	-31.7490	0.0712	0.2728	0.7516	3.2267	32.0131
		0.01%	1.1609	-0.1053	-0.3328	-0.8903	-3.4952	-32.6460	0.1234	0.3549	0.8988	3.5223	32.9009
		0.1%	9.9717	-0.8173	-1.1517	-2.2527	-6.2214	-40.8276	0.8203	1.1594	2.2570	6.2363	41.0226
TTS	5-sec	0%	0.1580	-0.0145	-0.2406	-0.7342	-3.1799	-31.6204	0.0424	0.2415	0.7436	3.2414	31.9329
		0.001%	0.2583	-0.0172	-0.2485	-0.7473	-3.2063	-31.6979	0.0434	0.2494	0.7564	3.2672	32.0093
		0.01%	1.1594	-0.0413	-0.3202	-0.8666	-3.4452	-32.4135	0.0576	0.3210	0.8744	3.5015	32.7163
		0.1%	9.9994	-0.2655	-0.9856	-1.9744	-5.6616	-39.0571	0.2689	0.9860	1.9775	5.6943	39.2982
	10-sec	0%	0.1581	-0.0119	-0.2265	-0.7183	-3.1534	-31.5369	0.0526	0.2332	0.7212	3.2016	31.8366
		0.001%	0.2584	-0.0150	-0.2355	-0.7333	-3.1832	-31.6278	0.0532	0.2420	0.7361	3.2309	31.9265
		0.01%	1.1597	-0.0432	-0.3163	-0.8679	-3.4518	-32.4340	0.0672	0.3212	0.8701	3.4954	32.7237
		0.1%	9.9805	-0.3034	-1.0619	-2.1114	-5.9387	-39.8976	0.3084	1.0636	2.1122	5.9628	40.1230
	20-sec	0%	0.1582	-0.0007	-0.2032	-0.6709	-3.0582	-31.2801	0.0641	0.2279	0.6732	3.0863	31.5484
		0.001%	0.2587	-0.0042	-0.2134	-0.6879	-3.0922	-31.3822	0.0642	0.2372	0.6902	3.1200	31.6493
		0.01%	1.1609	-0.0347	-0.3038	-0.8372	-3.3914	-32.2790	0.0734	0.3210	0.8390	3.4164	32.5371
		0.1%	9.9717	-0.3143	-1.1303	-2.2020	-6.1202	-40.4618	0.3219	1.1356	2.2029	6.1332	40.6588

Table C.2: ME and RMSE for NSR estimate M1A using $\overline{RV}_K$ under the CTS and TTS schemes

Scheme	Sparsity	NSR	True NSR	ME ($\times 10^4$)					RMSE ($\times 10^4$)				
			($\times 10^4$)	1-min	3-min	5-min	10-min	30-min	1-min	3-min	5-min	10-min	30-min
CTS	5-sec	0%	0.1580	-0.0184	-0.2468	-0.7461	-3.2049	-31.7167	0.0438	0.2497	0.7564	3.2661	32.0281
		0.001%	0.2583	-0.0211	-0.2547	-0.7591	-3.2314	-31.7939	0.0450	0.2576	0.7691	3.2920	32.1043
		0.01%	1.1594	-0.0449	-0.3263	-0.8781	-3.4696	-32.5084	0.0603	0.3285	0.8867	3.5257	32.8103
		0.1%	9.9994	-0.2669	-0.9913	-1.9853	-5.6853	-39.1497	0.2706	0.9923	1.9889	5.7179	39.3905
	10-sec	0%	0.1581	-0.0187	-0.2451	-0.7449	-3.2022	-31.7235	0.0544	0.2556	0.7502	3.2485	32.0200
		0.001%	0.2584	-0.0223	-0.2541	-0.7601	-3.2322	-31.8144	0.0555	0.2643	0.7653	3.2781	32.1099
		0.01%	1.1597	-0.0543	-0.3350	-0.8950	-3.5016	-32.6199	0.0746	0.3430	0.8994	3.5438	32.9066
		0.1%	9.9805	-0.3532	-1.0813	-2.1387	-5.9895	-40.0791	0.3576	1.0843	2.1409	6.0131	40.3029
	20-sec	0%	0.1582	-0.0263	-0.2357	-0.7268	-3.1648	-31.6494	0.0686	0.2651	0.7369	3.1951	31.9146
		0.001%	0.2587	-0.0361	-0.2457	-0.7434	-3.1987	-31.7511	0.0730	0.2741	0.7533	3.2286	32.0150
		0.01%	1.1609	-0.1226	-0.3347	-0.8922	-3.4971	-32.6481	0.1381	0.3563	0.9006	3.5242	32.9029
		0.1%	9.9717	-0.9524	-1.1539	-2.2545	-6.2233	-40.8297	0.9548	1.1614	2.2587	6.2382	41.0247
TTS	5-sec	0%	0.1580	-0.0151	-0.2411	-0.7347	-3.1805	-31.6210	0.0423	0.2420	0.7441	3.2420	31.9335
		0.001%	0.2583	-0.0178	-0.2490	-0.7478	-3.2069	-31.6985	0.0433	0.2499	0.7570	3.2678	32.0099
		0.01%	1.1594	-0.0419	-0.3206	-0.8672	-3.4457	-32.4141	0.0578	0.3214	0.8749	3.5020	32.7169
		0.1%	9.9994	-0.2660	-0.9857	-1.9749	-5.6621	-39.0574	0.2692	0.9861	1.9779	5.6947	39.2985
	10-sec	0%	0.1581	-0.0140	-0.2286	-0.7204	-3.1554	-31.5390	0.0518	0.2350	0.7232	3.2036	31.8387
		0.001%	0.2584	-0.0171	-0.2376	-0.7354	-3.1852	-31.6299	0.0527	0.2438	0.7382	3.2329	31.9286
		0.01%	1.1597	-0.0453	-0.3183	-0.8699	-3.4537	-32.4361	0.0676	0.3231	0.8722	3.4974	32.7258
		0.1%	9.9805	-0.3055	-1.0639	-2.1137	-5.9406	-39.8998	0.3100	1.0656	2.1145	5.9646	40.1252
	20-sec	0%	0.1582	-0.0086	-0.2111	-0.6786	-3.0662	-31.2884	0.0605	0.2335	0.6807	3.0943	31.5566
		0.001%	0.2587	-0.0120	-0.2214	-0.6956	-3.1002	-31.3904	0.0611	0.2428	0.6977	3.1279	31.6574
		0.01%	1.1609	-0.0423	-0.3117	-0.8449	-3.3993	-32.2872	0.0735	0.3274	0.8466	3.4243	32.5451
		0.1%	9.9717	-0.3221	-1.1380	-2.2093	-6.1277	-40.4696	0.3282	1.1428	2.2100	6.1407	40.6666

Table C.3: ME and RMSE for NSR estimate M2 using $\overline{RV}_K$ under the CTS and TTS schemes

Scheme	Sparsity	NSR	True NSR	ME ($\times 10^4$)					RMSE ($\times 10^4$)				
			($\times 10^4$)	2-min	3-min	5-min	10-min	30-min	2-min	3-min	5-min	10-min	30-min
CTS	5-sec	0%	0.1580	0.0643	0.0967	0.1629	0.3283	0.9923	0.0762	0.1093	0.1761	0.3412	1.0044
		0.001%	0.2583	0.0643	0.0967	0.1629	0.3283	0.9920	0.0763	0.1094	0.1760	0.3412	1.0042
		0.01%	1.1594	0.0644	0.0967	0.1627	0.3278	0.9903	0.0767	0.1098	0.1761	0.3408	1.0025
		0.1%	9.9994	0.0641	0.0955	0.1599	0.3224	0.9733	0.0809	0.1113	0.1750	0.3361	0.9855
	10-sec	0%	0.1581	0.0636	0.0962	0.1630	0.3283	0.9928	0.0759	0.1092	0.1763	0.3412	1.0050
		0.001%	0.2584	0.0632	0.0960	0.1628	0.3281	0.9924	0.0756	0.1088	0.1760	0.3409	1.0046
		0.01%	1.1597	0.0590	0.0926	0.1599	0.3251	0.9881	0.0728	0.1062	0.1734	0.3380	1.0003
		0.1%	9.9805	0.0154	0.0588	0.1302	0.2951	0.9468	0.0581	0.0839	0.1487	0.3099	0.9594
	20-sec	0%	0.1582	0.0462	0.0824	0.1509	0.3175	0.9840	0.0615	0.0968	0.1647	0.3308	0.9964
		0.001%	0.2587	0.0365	0.0749	0.1446	0.3119	0.9786	0.0547	0.0904	0.1589	0.3254	0.9911
		0.01%	1.1609	-0.0509	0.0076	0.0884	0.2617	0.9303	0.0658	0.0516	0.1104	0.2775	0.9433
		0.1%	9.9717	-0.9062	-0.6513	-0.4630	-0.2322	0.4561	0.9081	0.6543	0.4686	0.2504	0.4814
TTS	5-sec	0%	0.1580	0.0667	0.0995	0.1655	0.3324	1.0028	0.0786	0.1122	0.1785	0.3452	1.0149
		0.001%	0.2583	0.0666	0.0993	0.1654	0.3322	1.0025	0.0785	0.1121	0.1784	0.3451	1.0146
		0.01%	1.1594	0.0664	0.0993	0.1651	0.3317	1.0006	0.0784	0.1121	0.1782	0.3445	1.0128
		0.1%	9.9994	0.0644	0.0975	0.1619	0.3257	0.9831	0.0793	0.1116	0.1757	0.3387	0.9951
	10-sec	0%	0.1581	0.0680	0.1021	0.1724	0.3453	1.0381	0.0809	0.1155	0.1858	0.3587	1.0509
		0.001%	0.2584	0.0682	0.1022	0.1724	0.3452	1.0380	0.0810	0.1155	0.1858	0.3586	1.0507
		0.01%	1.1597	0.0680	0.1019	0.1719	0.3444	1.0357	0.0813	0.1155	0.1854	0.3578	1.0484
		0.1%	9.9805	0.0666	0.0998	0.1682	0.3374	1.0155	0.0863	0.1162	0.1831	0.3514	1.0282
	20-sec	0%	0.1582	0.0683	0.1020	0.1698	0.3396	1.0207	0.0822	0.1158	0.1836	0.3528	1.0333
		0.001%	0.2587	0.0681	0.1019	0.1697	0.3394	1.0204	0.0822	0.1156	0.1835	0.3527	1.0330
		0.01%	1.1609	0.0681	0.1017	0.1693	0.3387	1.0181	0.0830	0.1159	0.1834	0.3521	1.0307
		0.1%	9.9717	0.0673	0.0997	0.1659	0.3317	0.9966	0.0932	0.1189	0.1825	0.3461	1.0093

Table C.4: ME and RMSE for NSR estimate M2A using $\overline{RV}_K$ under the CTS and TTS schemes

Scheme	Sparsity	NSR	True NSR	ME ($\times 10^4$)					RMSE ($\times 10^4$)				
			($\times 10^4$)	2-min	3-min	5-min	10-min	30-min	2-min	3-min	5-min	10-min	30-min
CTS	5-sec	0%	0.1580	0.0624	0.0948	0.1611	0.3265	0.9905	0.0732	0.1069	0.1739	0.3392	1.0026
		0.001%	0.2583	0.0624	0.0949	0.1611	0.3265	0.9902	0.0732	0.1069	0.1739	0.3392	1.0023
		0.01%	1.1594	0.0623	0.0947	0.1608	0.3259	0.9884	0.0733	0.1070	0.1738	0.3387	1.0005
		0.1%	9.9994	0.0604	0.0924	0.1571	0.3197	0.9707	0.0744	0.1068	0.1715	0.3331	0.9828
	10-sec	0%	0.1581	0.0606	0.0936	0.1605	0.3259	0.9903	0.0718	0.1060	0.1734	0.3386	1.0025
		0.001%	0.2584	0.0595	0.0928	0.1598	0.3252	0.9896	0.0708	0.1051	0.1727	0.3380	1.0018
		0.01%	1.1597	0.0494	0.0850	0.1532	0.3190	0.9822	0.0631	0.0986	0.1667	0.3319	0.9945
		0.1%	9.9805	-0.0520	0.0081	0.0875	0.2570	0.9112	0.0717	0.0567	0.1117	0.2734	0.9242
	20-sec	0%	0.1582	0.0403	0.0775	0.1465	0.3134	0.9801	0.0557	0.0918	0.1603	0.3267	0.9925
		0.001%	0.2587	0.0280	0.0680	0.1385	0.3063	0.9733	0.0476	0.0839	0.1530	0.3198	0.9858
		0.01%	1.1609	-0.0828	-0.0174	0.0673	0.2427	0.9126	0.0916	0.0523	0.0936	0.2594	0.9258
		0.1%	9.9717	-1.1674	-0.8524	-0.6310	-0.3818	0.3167	1.1685	0.8543	0.6348	0.3927	0.3519
TTS	5-sec	0%	0.1580	0.0660	0.0988	0.1649	0.3318	1.0022	0.0776	0.1113	0.1778	0.3445	1.0143
		0.001%	0.2583	0.0658	0.0987	0.1648	0.3316	1.0019	0.0774	0.1112	0.1777	0.3444	1.0140
		0.01%	1.1594	0.0657	0.0986	0.1645	0.3311	1.0000	0.0774	0.1112	0.1775	0.3439	1.0122
		0.1%	9.9994	0.0642	0.0968	0.1614	0.3253	0.9826	0.0773	0.1104	0.1749	0.3382	0.9946
	10-sec	0%	0.1581	0.0656	0.1000	0.1702	0.3432	1.0360	0.0770	0.1126	0.1833	0.3564	1.0487
		0.001%	0.2584	0.0658	0.1001	0.1703	0.3431	1.0359	0.0771	0.1126	0.1833	0.3563	1.0485
		0.01%	1.1597	0.0657	0.0998	0.1698	0.3423	1.0336	0.0773	0.1126	0.1830	0.3556	1.0463
		0.1%	9.9805	0.0643	0.0977	0.1662	0.3353	1.0134	0.0792	0.1123	0.1803	0.3490	1.0260
	20-sec	0%	0.1582	0.0603	0.0941	0.1618	0.3317	1.0128	0.0696	0.1052	0.1742	0.3442	1.0252
		0.001%	0.2587	0.0604	0.0941	0.1618	0.3316	1.0126	0.0697	0.1051	0.1742	0.3442	1.0250
		0.01%	1.1609	0.0608	0.0942	0.1618	0.3312	1.0106	0.0703	0.1054	0.1742	0.3438	1.0229
		0.1%	9.9717	0.0595	0.0919	0.1580	0.3239	0.9888	0.0728	0.1056	0.1718	0.3370	1.0011

Table C.5: Empirical NSR estimates of 40 stocks

Stock	M1	M1A	M2	M2A	M3	Stock	M1	M1A	M2	M2A	M3
XOM	-94.29	-94.29	3.28	3.27	25.79	WMT	-94.49	-94.49	3.22	3.21	17.51
GE	-69.94	-69.94	2.76	2.75	21.80	CVX	-76.54	-76.54	2.79	2.78	11.85
IBM	-93.63	-93.64	3.22	3.21	23.48	JNJ	-96.01	-96.01	3.22	3.21	33.25
T	-97.82	-97.82	3.39	3.39	20.92	PG	-120.91	-120.90	4.43	4.42	17.40
JPM	-71.99	-71.99	2.35	2.35	11.22	WFC	-88.24	-88.24	2.85	2.84	18.44
KO	-93.60	-93.60	3.26	3.25	14.65	PFE	-85.93	-85.93	3.60	3.59	27.80
C	-46.01	-46.03	8.29	8.30	4.98	BAC	-47.22	-47.22	2.90	2.89	-2.36
SLB	-87.37	-87.37	3.01	3.01	14.12	MRK	-102.61	-102.61	3.36	3.35	16.70
PEP	-101.42	-101.42	3.32	3.31	17.25	VZ	-85.02	-85.02	3.04	3.03	10.83
COP	-81.14	-81.14	2.55	2.54	16.42	GS	-60.93	-60.93	2.00	1.99	-8.07
MCD	-67.26	-67.26	2.49	2.48	7.16	OXY	-68.65	-68.65	2.19	2.18	12.01
ABT	-89.10	-89.10	2.87	2.85	24.12	UTX	-77.10	-77.10	2.43	2.42	14.44
UPS	-92.42	-92.42	3.09	3.08	22.43	F	-54.46	-54.46	2.61	2.61	13.15
DIS	-70.95	-70.95	2.31	2.29	10.47	MMM	-125.09	-125.09	3.92	3.91	11.73
CAT	-55.09	-55.09	1.91	1.91	2.04	FCX	-61.73	-61.74	2.03	2.02	-0.70
USB	-75.92	-75.92	2.49	2.48	19.04	MO	-90.27	-90.28	3.52	3.51	33.06
AXP	-70.32	-70.32	2.34	2.33	2.75	BA	-75.59	-75.59	2.47	2.45	5.16
MDT	-75.77	-75.77	2.23	2.21	15.35	HD	-71.07	-71.07	2.39	2.38	11.48
CVS	-82.12	-82.12	2.72	2.70	15.76	EMC	-78.01	-78.00	2.67	2.65	20.94
HAL	-66.56	-66.55	2.23	2.22	6.65	PNC	-77.01	-77.02	2.40	2.39	14.28

Notes: NSR is calculated under the TTS scheme. We use $\overline{RV}_{KA^*}$ at 30-min sampling frequency for M1, M1A and M3 and at 1-min versus 30-min sampling frequencies for M2 and M2A, 2010-2013. The figures are after multiplying by 10^4.